

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 13, 2003, 23:27:52 ; Search time 3327 Seconds  
(without alignments)  
5009.064 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 5241  
Sequence: 1 MHILVHANYILLTLCPPRA.....MDPSAQISEALRLHMEAVM 1029

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 segs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 33308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus.p2n.model -DEV=xlh  
-O=/cgnr/\_1/USPRO.spool/US09929769/Unat\_07052003.161925-8044/app.query.fasta.1.1223  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi -LIST=45  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -ALIGN=15 -MODE=LOCAL  
-USER=US09929769.ecgn.1.1.2475 @unat.07052003.161925.8044 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-MARK\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST :  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estinu:\*  
5: em\_estlov:\*  
6: em\_estlpl:\*  
7: em\_estrio:\*  
8: em\_htc:\*  
9: gb\_estcl:\*  
10: gb\_estl2:\*  
11: gb\_htc:\*  
12: gb\_estl3:\*  
13: gb\_estl4:\*  
14: gb\_estl5:\*  
15: em\_estlun:\*  
16: em\_estlom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1543	29.4	981	9	AL518865
2	1461.5	27.9	959	14	B0071036
3	1431.5	27.3	951	14	B09440921
4	1411	26.9	971	13	BM548869
5	1379	26.3	896	14	B0878711
6	1374.5	26.2	896	14	B0686838
7	1327.5	25.3	1054	14	BM810364
8	1313	25.1	831	12	BG761156
9	1295	24.7	1098	13	B1912745
10	1292	24.7	823	12	BG828224
11	1278.5	24.4	800	12	BG825357
12	1278	24.4	858	12	BG768219
13	1218	23.2	897	14	B0953580
14	1217	23.2	728	12	BG468191
15	1206	23.0	899	12	BE792741
16	1205	23.0	994	12	BE530701
17	1203.5	23.0	730	9	AL042953
18	1199	22.9	766	13	B1835118
19	1195	22.8	729	13	B1554130
20	1190.5	22.7	859	12	BF688825
21	1190	22.7	840	12	BG765306
22	1182	22.6	810	13	B1259297
23	1172.5	22.4	1000	12	BG122941
24	1149	21.9	970	14	B0918808
25	1137	21.7	752	13	B1259197
26	1124	21.4	872	12	BG424981
27	1118.5	21.3	923	9	AL526487
28	1111.5	21.2	1001	12	BG749177
29	1111.5	21.2	1045	10	BE561214
30	1108	21.1	706	13	BM011375
31	1104	21.1	923	9	AL526445
32	1103.5	21.1	833	12	BE729570
33	1100.5	21.0	673	10	BE535502
34	1099.5	21.0	736	12	BG822646
35	1093.5	20.9	939	12	BF342499
36	1089	20.8	799	13	B1091872
37	1087	20.7	732	12	BG768009
38	1084.5	20.7	786	14	BQ215952
39	1074.5	20.5	825	12	BG753046
40	1073	20.5	738	13	BM018791
41	1068	20.4	684	12	BG774322
42	1066	20.3	748	13	BM050887
43	1059	20.2	766	10	BE410337
44	1057.5	20.2	1013	14	B0715615
45	1057.5	20.2	1018	12	BG328409

ALIGNMENTS

RESULT 1									
AL518865	AL518865	981 bp	mRNA	linear	EST 13-FEB-2001				
LOCUS	AL518865	LTI_NFL011.NBC1	Homo sapiens	cDNA clone	CSDBA011MY19 5				
DEFINITION	AL518865	prime, mRNA sequence.							
ACCESSION	AL518865								
VERSION	AL518865.1	GI:12782358							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
REFERENCE	1 (bases 1 to 981)								
AUTHORS	Li, W.B., Gruber, C.J., Jesse, J. and Polayes, D.								
TITLE	Full-length cDNA libraries and normalization								

JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr

FEATURES  
 source  
 1. 981  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS04011YM19"  
 /clone\_1ib="LTI\_NFL011\_NBC1"  
 /sex="male"  
 /tissue\_type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: Brain; Vector: PCWSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 173 a 328 c 303 g 164 t 13 others

ORIGIN

Alignment Scores:  
 Pred. No.: 7, 97e-145  
 Score: 1543.00  
 Percent Similarity: 94.17%  
 Best Local Similarity: 93.25%  
 Query Match: 29,44%  
 DB: 9  
 Gaps: 0

US-09-929-769-7 (1-1029) x AL518865 (1-981)

QY 614 SerValArgLysValThrGlnHisLeuSerGlyCysIleGlnGlnTrpGlyAspSerVal 633  
 |||||  
 DB 3 AGTGTGAGGAGGCTGACGAGGACCTGTGTGATGCTGATGAGGAGGAGGAGGCTG 62

QY 634 LeuGlyArgArgCysArgAspLeuLeuLeuGlnLeuTrpLeuGlnArgProGlnLeuArg 653  
 |||||  
 DB 63 CTGGGAGGAGGCTGCGAGGACCTTCTCTGCGGCTACTTACAGCGCGGAGGAGGCTG 122

QY 654 ValProValProGluValLeuLeuHisSerGluGlyAlaIleSerSerSerValCysLys 673  
 |||||  
 DB 123 GTGCCGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182

QY 674 LeuAspGlyLeuLeuHisArgPheIleThrLeuLeuAlaAspThrSerAspSerArgAla 693  
 |||||  
 DB 183 CKGAGGAGGACTCATCCACCCTTATCCACCCTCTTCCGAGACACAGGAGGAGGAGG 242

QY 694 LeuGlnAsnArgGlyAlaAspAlaSerMetAlaCysArgLysLeuAlaValAlaHisPro 713  
 |||||  
 DB 243 TTGGAGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 302

QY 714 LeuLeuLeuLeuArgHisLeuProMetIleAlaIleLeuLeuHisGlyArgThrHisLeu 733  
 |||||  
 DB 303 CTGCTGTGCTGACGAGGACCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362

QY 734 AsnPheGlnGlnPheArgGlnGlnHisLeuSerCysPheLeuHisValLeuGlyLeu 753  
 |||||  
 DB 363 AACCTGACGAGGAGTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 422

QY 754 LeuGlnLeuLeuGlnProHisValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCys 773  
 |||||  
 DB 423 CTGAGAGGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 482

QY 774 LeuLeuSerPheIleArgLeuLeuLeuAsnTrpArgLysSerSerArgHisLeuAlaIle 793  
 |||||  
 DB 483 CTCTGTCTTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 542

QY 794 PheIleAsnLysPheValGlnPheIleHisLysTrpIleThrTrpAsnAlaProAlaIle 813  
 |||||  
 DB 543 TTGATCAACAAGATTGTGGACGATTATCATTAAGTACATTACCTACATGCCCCAGCAGCC 602

QY 814 IleSerPheLeuGlnLysHisAlaAspProLeuHisAspLeuSerPheAspAsnSerAsp 833  
 |||||  
 DB 603 ATCTCTTCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 662

QY 834 LeuValMetLeuLysSerLeuLeuAlaGlyLeuSerLeuProSerArgAspAspArgThr 853  
 |||||  
 DB 663 CTGCTGATGCTGATATCTCTCTGAGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722

QY 854 AspArgLysLeuAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 873  
 |||||  
 DB 723 GACGAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 782

QY 874 SerValSerLeuPheThrProLeuThrAlaAlaGluMetLapProLysArgLeu 893  
 |||||  
 DB 783 AGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 842

QY 894 SerArgLysGlnThrValGluAspLeuLeuGlnValLeuSerAspIleAspGluMetSer 913  
 |||||  
 DB 843 TCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 902

QY 914 ArgArgArgProGlnLysLeuSerPhePheSerThrAsnLeuGlnArgLysMetSerSer 933  
 |||||  
 DB 903 CGCGGAGAGCCCGAGACCTGAGCTTCTCTGACCACTGCGAGCGGCTGAGAGAGCYCG 962

QY 934 AlaGlnGlnCysCysArg 939  
 |||||  
 DB 963 GCCGAGGAGGTGTGCGCG 980

RESULT 2  
 LOCUS B0071036 959 bp mRNA linear EST 02-APR-2002  
 DEFINITION AGENCOURT\_6853141 NIH\_MGC\_47 Homo sapiens cDNA clone IMAGE:5927385  
 5', mRNA sequence.  
 B0071036  
 B0071036.1 GI:19900082  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 959)  
 NIH-MGC http://mhc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 DNA sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LICM2102 row: d column: 10  
 High quality sequence stop: 705.  
 Location/Qualifiers  
 1. 959  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5927385"  
 /clone\_1ib="NIH\_MGC\_47"  
 /tissue\_type="neuroblastoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Brain; Vector: pORF7; Site: 1: XhoI, Site: 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(C). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in





```

|||||
Db 123 AAACCAAGAGCAGCCAGAGCCATAGAGCCAGGGCGGATTCGGGGAGCCAG 182
    182 LeuArgValLeuGlyProGluAspAlaLeuAlaGlyMetPheLeuGlnIlePheProLeu 201
    183 CTCGGGCTGCTGGCCCTAGAGACAGACTGGCTGCGATCTCTCCAGATTTTCCGCTC 242
    202 SerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 221
    243 AGCCGAGACCTCGTGGGAGAGCTCCAGTCCCGCCCGCTGGCCCTCGCCCTGACAGAG 302
    222 AlaLeuGlyGlnIleLeuAlaArgValValGlnGlySerProGluValProGlyIleThr 241
    303 GCCCTGGCCAGAGAGCTGGCCGCTGCTCAGAGGACGCCAGAGTGGCCGGGCAATCAG 362
    242 ValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyAlaLeuVal 261
    363 GTCCGCTCTCTGACAGCCCTCGCCACCTCTCTAGCTCCCCACACAGCGGCTGCTGCTG 422
    262 MetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuGlyGlnTyr 281
    423 ATGTCATGACACCTAGCCACTTCTGCGCTGCGCCCTGCGCTGCGCCAGCTCTCCAGTAC 482
    282 GluArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValLeuGln 301
    483 CAGCGCTGTGTGCCACAGACAGCGGCTTCTCTCTCTCTCTGAGAGGTCTCTGACAG 542
    302 MetLeuGlnTrpLeuAspSerProGlyValGlnGlyGlyProLeuArgAlaGlnLeuArg 321
    543 ATGCTGACAGTGGCTGGACAGCCCTGGGGGTGAGAGCGCGCCCTCGCGGACAGACTCAG 602
    322 MetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValAlaGlyGlyLeuLeu 341
    603 ATGCTTGGCCAGC----- 614
    342 ArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValValSerSerThrValArg 361
    615 ---CTGGCCGAGGCCCTGGCTTCCGCTCAGAGACTGGAGGTGCTCAGCTCCACCGTCCGT 671
    362 AlaValIleLeuAlaThrLeuArgSerGlyGlnGlnCysSerValGluProAspLeuIleSer 381
    672 GCCGTATCCCCACCTGAGAGTCTGGGGAGACAGTGCAGCTGGAGCCGAGCATGATCAGC 731
    382 LysValLeuGlnGlnLeu---IleGluValArgSerProHisLeuGlnLeuLeuThr 401
    732 AAAGTCTCTCCAGGGGGCTGATCAAGGTGAGTCCCGCCACCTGGAGAGCTGCTGACTG 791
    401 IapPheSerAlaThrAlaAspAlaAlaSerProPheProAlaCysLysPro-ValVal 420
    792 CATTTCTCTGTGCACATGGCGATGCTGCTCCCGCTTCCAGCTGTAACCCNCTTGGC 851
    421 ValValSerSerLeuLeuLeuGlnGlnGlnGlnProLeuAlaGlyGlyLysProGlyAla 440
    852 GGGGGAACACTCCTGCTGCTGCAAGAGAGAGAGACCCCTGCTGGGGGGAACCCGGGTTG 911
    441 -AspGlyGly-SerLeuGlnAlaValArgLeuGly 451
    912 GGACGGTGGCAACCTGGAAGCCCTGGCTGGGG 946

```

```

RESULT 4
LOCUS      BM548869          971 bp      mRNA      linear      EST 20-FEB-2002
DEFINITION AGENCOUNT_6543289 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5740983
ACCESSION  BM548869
VERSION     BM548869.1 GI:18783808
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 971)
AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.

```

```

TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph. D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12756 row: m column: 16
            High quality sequence stop: 636.
FEATURES   location/Qualifiers
            source
            1. 971
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5740983"
                /clone_lib="NIH_MGC_88"
                /tissue_type="duodenal adenocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: small intestine; Vector: pCMV-SPORT6;
                Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally;
                oligo-dr primed. Average insert size 1.767 kb. Library
                enriched for full-length clones and constructed by Life
                Technologies. Note: this is a NIH_MGC library."
BASE COUNT 150 a 332 c 331 g 158 t
ORIGIN
Alignment Scores:
Pred. No.: 1 71e-131      Length: 971
Score: 1411.00           Matches: 277
Percent Similarity: 97.89%      Conservative: 2
Best Local Similarity: 97.19%      Mismatches: 4
Query Match: 26.92%           Indels: 2
DB: 13                      Gaps: 0
US-09-929-769-7 (1-1029) x BM548869 (1-971)
    410 AlaserProPheProAlaCysLysProValValValSerSerLeuLeuGlnGln 429
    3 GCCTCCCCGTTTCCAGCTGTAAGCCCGCTGTGGTGTGAGCTCCCTGCTGACAGAG 62
    430 GlnGlnProLeuAlaGlyGlyLysProGlyAlaAspAlaGlySerLeuGlnAlaValArg 449
    63 GAGAGGCCCTCTGGCTGGGGGGAAGCCGGGTGGGACCGTGGACGCTGGAGGCCGTGG 122
    450 LeuGlyProSerSerGlyLeuLeuValAspTrpLeuGlnMetLeuAspProGluValVal 469
    123 CTGGGGCCCTCGTACAGCCCTCCTAGTGTGACTGCTGGAATGCGTGGACCCCGAGCTGCT 182
    470 SerSerCysProAspLeuGlnLeuArgLeuLeuPheSerArgArgLysGlyGlyGln 489
    183 AGCAGTGCCTCCGACCTGACAGCTCCTCTTCTCCCGGAGAGAGGCAAGCTCAG 242
    490 AlaGlnAlaProSerPheArgProTyrLeuLeuThrLeuPheThrHisGlnSerSerTrp 509
    243 GCCCAGGTGGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 302
    510 ProThrLeuHisGlnCysIleArgValLeuLeuGlnGlyLysSerArgGlnArgPheAsp 529
    303 CCCACACTGCACCACTGATCCAGTCTCTCTGGGCAAGACCGGGAACAGAGTTGAC 362
    363 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 422
    550 ArgAspGlnArgThrProGlnLysArgArgGlnGlnLeuValLeuArgValGlnGlyPro 569
    423 CGGACACAGGACACCCCGGAGAGCGGGGAGAGAGTGTGCTGCTGCGGCTCAGGGCCG 482
    570 GlnLeuIleSerLeuValGlnLeuIleLeuAlaGlnAlaGlnArgArgSerGlnAspGly 589

```

Db 483 GAGCTCATGACCTGGTGGACCTGATCTGGCCGAGGGGAGACCGGAGCGAGCGG 542  
 QY 590 AsptHra1a1aCysSerLeu1leGln1a1aArgLeuProLeuLeuSerCysCys 609  
 Db 543 GACACAGCGCCCTGACGCTTATCCAGGCGGCGCTGCTCTTACGCTGCTGCT 602  
 QY 610 G1yAspAspGluSerValArgLysValThrGlnHisLeuSerGlyCys1leGlnGlnTrp 629  
 Db 603 GGGGAGCATGAGAGTGTAGAGAGGTGACGAGGACGCTGTACGCTGCATCCAGAGTGG 662  
 QY 630 G1yAspSerValLeuGlyArgArgCysArgAspLeuLeuGlnLeuTrpLeuGlnArg 649  
 Db 663 GGAGACACCGCTGCGGAGGCGGCGGACCTTCTCTGACCTTACCTACAGCGG 722  
 QY 650 ProGluLeuArgVal-ProValProGluValLeuLeuHisSerGluGlyAla1a1aSerSe 669  
 Db 723 CCGGAGCTGGCGGTCGCCGCTGAGTCCCTACTGTCAGACGAGGAGGCGGCGCAGAG 782  
 QY 669 rSerValCysLysLeuAspGlyLeu1leHisArgPhe1leThr-LeuLeuAlaAspThrS 689  
 Db 783 CAGGCTGTCCAGCTGAGGAGGAGCTATCCACCGGCTTATCCAGGCTCTTGGCGGACCCA 842  
 QY 689 erAspSerArg 692  
 Db 843 CCGACTCCCGG 853  
 RESULT 5  
 B0878711  
 LOCUS B0878711 896 bp mRNA linear EST 16-AUG-2002  
 DEFINITION AGNCOURT\_8118299 Lupski\_dorsal\_root\_ganglion Homo sapiens CDNA  
 clone IMAGE:6181163 5', mRNA sequence.  
 B0878711  
 VERSION B0878711.1 GI:22270719  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 896)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: rgsb@femail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/MLN at:  
 http://image.lnl.gov  
 Plate: LRAM13565 row: b column: 12  
 High quality sequence stop: 530.  
 Location/Qualifiers  
 1. 896  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6181163"  
 /clone\_lib="Lupski\_dorsal\_root\_ganglion"  
 /sex="male"  
 /tissue\_type="dorsal root ganglia"  
 /dev\_stage="adult, 36 yr"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); site\_1:  
 NotI, site\_2: SalI; CDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TCGACCCACGCGCTCG-3' and  
 5'-GACTAGTCTGATCGGAGCGGCCCTT(15)-3', size selected >  
 1 kb for average insert length 1.7 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
 College of Medicine) and is available through Life  
 Technologies."

BASE COUNT 161 a 312 c 253 g 164 t 6 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2.54e-128 Length: 896  
 Score: 1379.00 Matches: 285  
 Percent Similarity: 95.67% Conservative: 12  
 Best Local Similarity: 95.00% Mismatches: 12  
 Query Match: 26.31% Indels: 3  
 DB: 14 Gaps: 0  
 US-09-929-769-7 (1-1029) x B0878711 (1-896)  
 QY 664 G1uG1yAla1a1aSerSerSerValCysLysLeuAspGlyLeu1leHisArgPhe1leThr 683  
 Db 1 GAGGAGGCTGGCAGACGACGAGCTGTGCAAGCTGGAGCGAGACTATGACCGCTTATCCAG 60  
 QY 684 LeuLeuAlaAspThrSerAspSerArgAlaLeuGluAsnArgGlyAlaAspAlaSerMet 703  
 Db 61 CTCCTGGGAGACACGAGGAGCTCCGGGCGTTGGAGAACCGAGGGCGGATGCCAGCATG 120  
 QY 704 AlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuArgHisLeuProMetIle 723  
 Db 121 GCGTCCGGAAGCTGGCGGTGGCGGACCGGCTGTCTGTCTGAGCACCCTGCCCATGATC 180  
 QY 724 AlaAlaLeuLeuHisGlyArgThrHisLeuAsnPheGlnGluPheArgGlnGlnAsnHis 743  
 Db 181 GCGGCGCTCTCTACAGCGGCGGACCCACCTCAACTTCCAGAGATGTCGCGCAGAGAACAC 240  
 QY 744 LeuSerCysPheLeuHisValLeuGlyLeuLeuGluLeuGlnProHisValPheArg 763  
 Db 241 CTGAGCTCTCTCTCTCAGCTGCTGGGCTGTGGAGCTGTGACAGCTGTGACAGCTGTCTCCG 300  
 QY 764 SerGlnHisGlnGlyAlaLeuThrAspCysLeuLeuSerPhe1leArgLeuLeuLeuAsn 783  
 Db 301 AGCGAGACACGAGGCGGCTGTGGAGCTGCTTGTCTGTCTTATCGGCTGTCTGTAGAT 360  
 QY 784 TyrArgLysSerSerArgHisLeuAla1a1aPhe1leAsnLysPheValGlnPhe1leHis 803  
 Db 361 TACAGAGAGTCTCTCCGCGCATGTGGCTGCTTCATCAACAATTTGTGAGTATTCAT 420  
 QY 804 LysTyr1leThrThrTyrAsnAlaProAlaAla1leSerPheLeuGlnLysHisAlaAspPro 823  
 Db 421 AAGTACATTAACCTACAAAGCCCGACAGCCAGCTCTCTCTCCAGAACGCCACCGCCG 480  
 QY 824 LeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeuLeuLagly 843  
 Db 481 CTCACGACCTGTCTCTTGACACAGTGAACCTGTGATGCTGAATCCCTCTTCAGGG 540  
 QY 844 LeuSerLeuProSerArgAspAspArgThrAspArgGlyLeuAspGluGluGlyLysLys 863  
 Db 541 CTCACCTGTCCGACGAGGAGGAGACAGACGAGCCAGCGGCTGTGAGAAAG-GGCGAGGG 599  
 QY 864 GluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThrAla 883  
 Db 600 GAGAGCTAGCGGCTCTCTGACACAGTGAACCTGTGATGCTGAATCCCTCTTCAGGG 659  
 QY 884 AlaGluMetAlaProTyrMetLysArgLeuSerArgGlyGlnThrValGluAspLeuLeu 903  
 Db 660 GNGGAGATGGCCCTTACTATGAAAGCGCTTTCGAGNGCCAAAGCGTGGAGATCTGCTG 719  
 QY 904 GluValLeuSerAsp1leAspGluMetSerArgArgArgProGlu1leLeuSerPhePhe 923  
 Db 720 GAGGTTCGAGTGAATGAGAGAGATGTCCGCGGAGAGCCGAGATCTGTGATCTTCTC 779  
 QY 924 SerThrAsnLeuGlnArgLeuMetSerSerAlaGluGluCysCysArgAsnLeuAlaPhe 943  
 Db 780 TCGAACCAACCTGACCGCTGATGAGCTCCCGCGGAGATGTGTGGC-AACTCCCTCTC 838  
 QY 944 SerLeu-AlaLeuArgSerMetGlnAsnSerProSer1leAla1a1aPheLeuPro 962  
 Db 839 ACCTGTGACCTGTGCGCTTCATGACAGAACGCCACGACATTCGAGCCGCTTTTCTGCC 896

RESULT 6  
B0686838 896 bp mRNA linear EST 15-JUL-2002  
LOCUS AGENCOURT 8171634 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6252041  
DEFINITION 5', mRNA sequence.  
ACCESSION B0686838  
VERSION B0686838.1 GI:21812154  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 896)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-rt@mail.nih.gov](mailto:cgapbs-rt@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Place: LNCM2397 row: k column: 18  
High quality sequence stop: 683.  
Location/Qualifiers  
1..896  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6252041"  
/clone\_1ib="NIH\_MGC\_110"  
/tissue\_type="ductal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pORF7; Site\_1: XhoI;  
Site\_2: EcoRI; CDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(S). Library constructed  
by Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH-MGC library."

BASE COUNT 157 a 307 c 274 g 158 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7,23e-128 Length: 896  
Score: 1374.50 Matches: 281  
Percent Similarity: 96.27% Conservative: 3  
Best Local Similarity: 95.25% Mismatches: 7  
Query Match: 26.23% Indels: 5  
DB: 14 Gaps: 1

US-09-929-769-7 (1-1029) x B0686838 (1-896)

QY 562 LeuValLeuArgValGlnGlyProGluLeuLeuSerLeuValGluLeuLeuAlaGlu 581  
|||||  
DB 2 CTGGTCTGGGGGTCCAGGCGCCGAGCTCATCAGCGTGTGGAGTGCATCTGGCCGAG 61  
|||||

QY 582 AlaGluThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuLeuGlnAlaArgLeu 601  
|||||  
DB 62 GCGGACACGGGGAGCCAGGACGCGGACACAGCCGCTGCAGCTCATCCAGCGCGCTG 121  
|||||

QY 602 ProLeuLeuLeuSerCysCysGlyAspAspGluSerValArgLysValIrrGluHis 621  
|||||  
DB 122 CCCCTCTGCTCACTGCTCTCTGTGGGACGATGAGGTCTCAGAGAGGTGACGACAC 181  
|||||

QY 622 LeuSerGlyCysIleGlnGlnIrrGlyAspSerValLeuGlyArgCysArgAspLeu 641  
|||||  
DB 182 CTCTCAGGCTGATCCAGCAGGTGGGAGACAGCGTGTGGGACGCGCTCCGAGACCTT 241  
|||||

QY 642 LeuLeuGlnLeuThrLeuGlnIrrGlyProGluLeuArgValProValProGluValLeuLeu 661  
|||||

DB 242 CTCTGAGCTCTACCTACAGCGCGAGCTCCGGGTGCGCTGCTGAGTCTACTG 301  
|||||

QY 662 HisSerGluGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuLeuHisArgPhe 681  
|||||

DB 302 CACAGCGAAGGGCTGCGACAGCAGCGCTTGCAGGTGAGCGACTATCCACCGCTTC 361  
|||||

QY 682 IleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGluAsnArgGlyAlaAspAla 701  
|||||

DB 362 ATCAGCGCTCTTCCGACACAGCAGCTCCCGGCGCTTGAGAACGAGGCGGATGCC 421  
|||||

QY 702 SerMetAlaCysArgLysLeuAlaValAlaHisIrrProLeuLeuLeuAlaHisLeuPro 721  
|||||

DB 422 AGCATGGCTCCGCGAAGCTGGGGTGGCGCCAGCCGCTGCTGCTCAGGCACTGCCC 481  
|||||

QY 722 MetIleAlaLeuLeuHisGlyArgThrHisLeuAsnPheGlnIrrPheArgGlnGln 741  
|||||

DB 482 ATGATCGCGCGCTCTCGACGCGCCGACCCACTCACTCCAGAGTTCGCGACAG 541  
|||||

QY 742 AsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGluLeuGlnProHisVal 761  
|||||

DB 542 AACCACTGAGCTGCTCTCGACAGTGTGGCTGCTGAGCTGCTGAGCTGCGACCGCAG 601  
|||||

QY 762 PheArgSerGluHisGlnGlyAlaLeuThrAspCysLeuLeuSerPheIleArgLeuLeu 781  
|||||

DB 602 TTCGCGACGAGACACAGGGGGCTGTGGAGCTGCTTGTCTTCATCCGCTGCTG 661  
|||||

QY 782 LeuAsnTrpArgLysSerSerArgHisLeuAlaAlaPheIleAsnLysPheValGlnPhe 801  
|||||

DB 662 CTGATATACAGAAAGTCTCTCCCGCATCTGCTGCTCATCAACAAGTTTGTGCACTTC 721  
|||||

QY 802 IleHisLysTrpIleThrTrpAsnAlaProAlaAlaIleSerPheLeuGlnLysHisAla 821  
|||||

DB 722 ATCATATAGTACATTAATCTCAATGCCCCAGCAGCATCTCTCTCGCA-AAAACAGCC 780  
|||||

QY 822 AspProLeuHisAspLeuSerPheAsp-AsnSerAspLeu-ValMetLeuLys- -SerL 840  
|||||

DB 781 GACCCGCTCCAGCAGCTGCTTTAAACAACAGTGCATGAGTGTGAAATCCCTT 840  
|||||

QY 840 euleuAlaGlyLeuSer-LeuProSerArgAspArg-852  
|||||

DB 841 CCTTCGAGGGGTTCAGCTTGGCCCCGAGGAGGACCAACAG 879  
|||||

RESULT 7  
BM810364 1054 bp mRNA linear EST 05-MAR-2002  
LOCUS AGENCOURT 6580583 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5453467  
DEFINITION 5', mRNA sequence.  
ACCESSION BM810364  
VERSION BM810364.1 GI:19127187  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 1054)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-rt@mail.nih.gov](mailto:cgapbs-rt@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Place: LNCM1947 row: a column: 20  
High quality sequence stop: 632.  
Location/Qualifiers  
1..1054  
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:5453467"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: brain; Vector: pORF7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."
BASE COUNT      172 a      358 c      336 g      186 t      2 others
ORIGIN
Alignment Scores:
Pred. No.:      5,47e-123      Length:      1054
Score:          1327.50      Matches:      274
Percent Similarity: 94.20%      Conservative: 2
Best Local Similarity: 93.52%      Mismatches: 10
Query Match:    25.33%      Indels:      7
Db:             14      Gaps:      1
US-09-929-769-7 (1-1029) x BM810364 (1-1054)
Oy 454 SerGlyLeuLeuValAspTrpLeuLumMetLeuAspProGluValSerSerCysPro 473
Db 1 TCAGGCTTCCTAGTGTGAGCTGGGTAATGCTGAGACCCCGAGAGTGTGACGAGCTGCC 60
Oy 474 AspleuGlnLeuArgLeuLeuPheSerArgArgGlyGlyGlyGlnAlaGlnAlaPro 493
Db 61 GACCTGCAGCTCAGCTGCTCTTCCCGGAGAGGGAAGGCAAGTCAGGCCAGCTGCC 120
Oy 494 SerPheArgProTrpLeuLeuThrLeuPheThrHisGlnSerSerTrpProThrLeuHis 513
Db 121 TCGTTCCTCCCTTACCTCTGACCTCTTCACGATCAGTCAGTCGCGCCACTGTGCAC 180
Oy 514 GlnCysIleArgValLeuLeuGlyLysSerArgGlnGlnArgPheAspProSerLaser 533
Db 181 CAGTGCATCCGAGTCTCTGCGGCAAGACCGGGAACAGGTTGACCCCTCTGCTCT 240
Oy 534 LeuAspPheLeuTrpAlaCysIleHisValProArgIleTrpGlnGlyArgAspGlnArg 553
Db 241 CTGAGCTTCCTCTGGGCTGCATCATCTCTGCACTGCGAGGGGGGAGCAGCAGCC 300
Oy 554 ThrProGlnLysArgArgGlnGlnLeuValLeuArgValGlnGlyProGluLeuLaser 573
Db 301 ACCCGCGAAGCGCGGAGAGCTGTGCTGCGGCTCCAGGCGCCGAGCTCATCAGC 360
Oy 574 LeuValGlnLeuIleLeuAlaGlnAlaGlnArgSerGlnAspGlyAspPheTrpAlaHis 593
Db 361 CTGTGTGAGCTGATCTGTGCGGAGGAGGAGCGGAGCCAGGAGGAGGAGCAGCCGCG 420
Oy 594 CysSerLeuIleGlnAlaArgLeuProLeuLeuLeuSerCysCysGlyAspAspGln 613
Db 421 TGCAGCTCATCCAGCGCGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
Oy 614 SerValArgLysValThrGlnHisLeuSerGlyCysIleGlnGlnTrpGlyAspSerVal 633
Db 481 AGTGTGAGAGAGTACGAGGAGCAGCTGTGAGCTCATCCAGCTGGGAGACAGCGGTG 540
Oy 634 LeuGlyArgArgCysArgAspLeuLeuGlnLeuTrpLeuGlnArgProGluLeuArg 653
Db 541 CTGGGAGAGCGCTGCGGAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Oy 653 GVal-ProValProGluValLeuLeuHisSer-GlnGlyAlaAlaSerSerSerValCys 672
Db 601 GGTGCGCCGTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Oy 673 LysLeuAspGlyLeuIleHisArgPheIleThrLeuLeuAlaAspThrSerAspArg 692
Db 661 AAGCTGAGGAGCTATCATCCACGCTTTCACGCTTTCGCGAGACCGGACTCCCGG 720

```

```

Oy 693 AlaLeuGlnu-AsnArgGlyAlaAspAlaSerMetAlaCysArgLys-LeuAlaValAlaH 712
Db 721 GGTGTGAGAGAGCCGAGGCGGATGTCAGCTGCGGGAACCTGGCGGTGGCGC 780
Oy 712 ISProLeuLeuLeuLeuArgHisLeuPheMet---IleAlaAlaLeuLeuHisGlyArgT 731
Db 781 ACCGCTGCTGTGCTGCTGAGGCGCCCTGCCATTGATCGGGGGGCTCCGCGACGGCGCA 840
Oy 731 hrHis-LeuAsnPhcGlnGluPheArg 739
Db 841 CCCACCTTAATTTTCAGGAATTCGCG 867
RESUT 8
BG761156
LOCUS      831 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION 602717492P1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4841134 5',
            mRNA sequence.
ACCESSION  BG761156
VERSION    BG761156.1 GI:14071809
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 831)
            NIH-MGC http://mgc.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: c9apbs-remail.nih.gov
            Tissue Procurement: ATCC/DCTD/DRP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: LICM1673 row: 9 column: 23
            High quality sequence stop: 778.
            Location/Qualifiers
                1..831
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4841134"
                /clone_lib="NIH_MGC_49"
                /tissue_type="melanocytic melanoma, high MDR (cell line)"
                /lab_host="DH10B (phage-resistant)"
                /Note="Organ: skin; Vector: pORF7; Site: 1: XhoI; Site: 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGCACGAG(G). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the Laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                II RT (Life Technologies). Note: this is a NIH_MGC
                library."
BASE COUNT      121 a      271 c      290 g      149 t
ORIGIN
Alignment Scores:
Pred. No.:      1.04e-121      Length:      831
Score:          1313.00      Matches:      267
Percent Similarity: 96.04%      Conservative: 0
Best Local Similarity: 96.04%      Mismatches: 10
Query Match:    25.05%      Indels:      3
Db:             12      Gaps:      1
US-09-929-769-7 (1-1029) x BG761156 (1-831)
Oy 332 ArgLeuSerAspValArgGlyGlyLeuLeuArgLeuAlaGlnAlaLeuAlaPheArgGln 351
Db 2 AGGCTCAGTGTGCTGAGGGGCTCTGCTGCTGCGCGAGGCGCTGCGCTTCCGTGAG 61

```

QY 352 AspleuGlulValSerSerThrValArgAlaValIleAlaThrLeuArgSerGlyLu 371  
 |||||  
 Db 62 GACCTGGAGGTGCTCAGCTCCACCGTCCGCGCATCCACCTGAGCTGGGAG 121  
 |||||  
 QY 372 GlnCysSerValGluProAspLeuIleSerValLeuGlnGlyLeuIleGluValArg 391  
 |||||  
 Db 122 CAGTGGAGCTGGAGCGGAGCTGATCAGCAAGATCCTCCAGGCGCTGATCGAGGTAG 181  
 |||||  
 QY 392 SerProHisLeuGluLuleuLeuThrAlaPheSerAlaThrAlaAspAlaLaser 411  
 |||||  
 Db 182 TCCCCCACTGGAGGAGAGCTGATGCTGCTTCTCTCCACTGGGATGCTGCTCC 241  
 |||||  
 QY 412 ProPheProAlaCysLysProValValValValSerSerLeuLeuGlnGluGlu 431  
 |||||  
 Db 242 CCGTTTCAGCTGTAAAGCCGCTGTGTGTAGAGCTCCCTGCTGCGAGAGAGAG 301  
 |||||  
 QY 432 ProLeuAlaGlyLysProGlyAlaAspGlyGlySerLeuGlnAlaValArgLeuGly 451  
 |||||  
 Db 302 CCGCTGGCTGGGGGAGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 361  
 |||||  
 QY 452 ProSerSerGlyLeuLeuValAspTrpLeuGluMetLeuAspProGluValLaser 471  
 |||||  
 Db 362 CCGCTGCAAGCTCTCTAGGAGCTGGCTGGAATGCTGGAACCCGAGGCTGAGCAGC 421  
 |||||  
 QY 472 CysProAspLeuGlnLeuArgLeuLeuPheSerArgArgLysGlyLysGlnAlaGln 491  
 |||||  
 Db 422 TGCCTCCGAGCTGAGCTGAGCTGCTCTTCCCGAGAGAGGAGGAGGAGGAGGAG 481  
 |||||  
 QY 492 ValProSerPheArgProGlyLeuLeuThrLeuPheThrHisGlnSerSerTrpThr 511  
 |||||  
 Db 482 GTGCGCTGCTGGTCCGCTCCCTACCTCTGAGCTGATCCCTTTCAGATGCTCAGCTG 541  
 |||||  
 QY 512 LeuHisGlnCysIleArgValLeuLeuGlyLysSerArgLysGlnArgPheAspProSer 531  
 |||||  
 Db 542 CTGCACACAGTGCATCCGAGCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 601  
 |||||  
 QY 532 AlaSerLeuAspPheLeuThrAlaCysIleHisValProArgIleTrpGlnIleArgAsp 551  
 |||||  
 Db 602 GCGTCTGTGAGACTCTCTGGGCTGCTGATCCATGCTCTGCTGCGAGGGGAGGAG 661  
 |||||  
 QY 552 GlnArgThrProGlnLysArgArgGlnGlnLeuValLeuArgValGlnGlnIleProGlnLeu 571  
 |||||  
 Db 662 CAGCGACACCCGAGAGAGCGGCGGAGAGAGCTGCTGCTGCGAGGCGGAGAGCTC 721  
 |||||  
 QY 572 IleSerLeuValGluLeuIleLeuAlaGlnAlaGluThrArgSerGlnAspGlyAspThr 591  
 |||||  
 Db 722 ATCAGGCTGTGGAGCTGATCTGCGCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAG 780  
 |||||  
 QY 592 AlaAlaCysSerLeuIleGlnAlaArgLeuProLeuLeuLeuSerCysCys 609  
 |||||  
 Db 781 GGGCG-TCGAGGCTCATGATGCCGATGCCCT--GCTCTCAGCTGCTCTGT 830  
 |||||

RESULT 9  
 BI912745 1098 bp mRNA linear EST 16-OCT-2001  
 LOCUS 603176136F1 NIH\_MGC\_121 Homo sapiens CDNA clone IMAGE:5240481 5'  
 DEFINITION mRNA sequence.  
 ACCESSION BI912745  
 VERSION BI912745.1 GI:16176988  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 1098)  
 AUTHORS NIH-MGC <http://imgc.ncl.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LLM11606 row: g column: 10  
 High quality sequence stop: 727.  
 Location/Qualifiers  
 1..1098  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5240481"  
 /clone\_lib="NIH\_MGC\_121"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMV-Sport6; Site:1: NotI;  
 Site:2: EcoRV (destroyed); RNA source anonymous pool of 3  
 fetal brains, female age 20 weeks, female age 24 weeks,  
 and male age 26 weeks. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.7 kb, insert size range  
 0.7-3.5 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics: tracking code 017. Note:  
 this is a NIH-MGC Library."

BASE COUNT 219 a 354 c 334 g 191 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,14e-119 Length: 1098  
 Score: 1295.00 Matches: 265  
 Percent Similarity: 93.738 Conservative: 4  
 Best Local Similarity: 92.338 Mismatches: 14  
 Query Match: 24,718 Indels: 5  
 DB: 13 Gaps: 0  
 US-09-929-769-7 (1-1029) x BI912745 (1-1098)  
 QY 470 SerSerCysProAspLeuGlnLeuArgLeuLeuPheSerArgArg-LysGlyLysGly 489  
 |||||  
 Db 2 AGCAGCTGCCCGACCGACCTGAGCTGCTCTTCCCGAGTGAAGGCGCAAGGTCA 61  
 |||||  
 QY 489 nAlaGlnValProSerPheArgProGlyLeuLeuThrLeuPheThrHisGlnSerSer 509  
 |||||  
 Db 62 GGGCCAGAGCTCCCTCGTCCGCTACCTCTGACCTTTTACGATAGTCCAGCTG 121  
 |||||  
 QY 509 pProThrLeuHisGlnCysIleArgValLeuLeuGlyLysSerArgGlnArgPheAs 529  
 |||||  
 Db 122 GCCCAGACTGCACAGAGTCCGAGCTGCTGGGAGAGAGCGGAGAGAGAGAGAGAG 181  
 |||||  
 QY 529 pProSerAlaSerLeuAspPheLeuThrAlaCysIleHisValProArgIleTrpGlnI 549  
 |||||  
 Db 182 CCGCTGCTGCTCTGTGAGCTCTCTGGGCTCTCATCTCATGCTCTGCGATCTGGAGG 241  
 |||||  
 QY 549 YArgAspGlnArgThrProGlnLysArgArgGlnGlnLeuValLeuArgValGlnGlyPr 569  
 |||||  
 Db 242 GCGGAGACAGCGACCCCGGAGAGCGGCGGAGAGAGCTGCTGGGGTCCA-66CCC 300  
 |||||  
 QY 569 OGluLeuIleSerLeuValGluLeuIleLeuAlaGlnAlaGluThrArgSerGlnAspG 589  
 |||||  
 Db 301 GAGAGCTCATGAGCTGGTGGAGCTGATCCCGCGGAGGCGGAGAGCGGAGCGAGCG 360  
 |||||  
 QY 589 YAspThrAlaAlaCysSerLeuIleGlnAlaArgLeuProLeuLeuLeuSerCysCys 609  
 |||||  
 Db 361 GGACACAGCGCGCTGAGCTCATCCAGGCGCGGCTGCTCTGCTGCTGCTGCTGCTG 420  
 |||||  
 QY 609 sGlyAspAspGlySerValArgLysValThrGlnHisLeuSerGlyCysIleGlnGlnTr 629  
 |||||  
 Db 421 TGGGAGAGATGATAGTCTCAGAAAGTGAAGGAGGAGACACTGACGCTGATCCAGCAGT 480  
 |||||  
 QY 629 pGlyAspSerValLeuGlnValArgArgCysArgAspLeuLeuGlnLeuThrArgLeuGlnAr 649  
 |||||  
 Db 481 GGGAGAGAGGCTGCTGGGAGGCGGCTGCGAGAGCTTCTCTCAGAGTCACTACAGAGCG 540  
 |||||

QY 649 gProgluLeuArgValProValProgluValLeuLeuHisSerGluGlyAlaAlaSeSe 669  
 Db 541 GCCGAGCTGGCGGGTCCGCTGAGCTCTACTGACAGGAGGCGGCTGCCAGCAG 600  
 QY 669 rSerValCysLysLeuAspGlyLeuIleHisArgPheIleIleHisLeuAlaAspThrSe 689  
 Db 601 CAGGCTGCGACGCGGAGCGAGCTATCCACCGCTTCATCCAGCTTCGCGAGACAG 660  
 QY 689 rAspSerArgAlaLeuGluAsnArgGlyAlaAspAlaSerMetAlaCysArgLysLeuAl 709  
 Db 661 CGACTCCCGGGCGCTGGAGAACCGAGGCGGATGCCAGATGGCTCCCGAGACTGCG 720  
 QY 709 aValAlaHisProLeuLeuLeuLeu-ArgHisLeuProMetIleAlaAla-LeuLeuH 728  
 Db 721 GGTGGCGAACCGCTGTGCTGCTCCAGGCGACCTGACCATGATGTCGAGAGCTCTTGC 780  
 QY 728 sGlyArgThrHisLeuAsnPheGlnGluPheArgGlnIleAsnHisLeuSerCysPheLe 748  
 Db 781 ACGGAGACCCCACTCACTCCAGAGCTCCGACAGAACCACTGAGCTGTATCCT 840  
 QY 748 uHisValLeuGly 752  
 Db 841 GTACTGTCATGGG 853

RESULT 10  
 LOCUS BG828224 823 bp mRNA linear EST 22-MAY-2001  
 DEFINITION 602747827F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4906590 5',  
 mRNA sequence.  
 ACCESSION BG828224  
 VERSION BG828224.1 GI:14175811  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 823)  
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: rgs@nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L1CM1809 row: 0 column: 07  
 High quality sequence stop: 819.  
 Location/Qualifiers  
 1. 823  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4906590"  
 /clone\_1id="NIH\_MGC\_17"  
 /tissue\_type="rhabdomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: muscle; Vector: pOT7; Site\_1: EcoRI;  
 Site\_2: XhoI; cDNA made by oligo-CT priming.  
 directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GCGACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 133 a 275 c 270 g 145 t  
 ORIGIN  
 Alignment Scores: 1.35e-119 Length: 823  
 Pred. NO.: 1292.00 Matches: 269  
 Score:

Percent Similarity: 97.83% Conservative: 1  
 Best Local Similarity: 97.46% Mismatches: 4  
 Query Match: 24.65% Indels: 6  
 DB: 12 Gaps: 0

US-09-929-769-7 (1-1029) x BG828224 (1-823)

QY 530 ProSerAlaSerLeuAspPheLeuTrpAlaCysIleHisValProArgIleTrpGlnGly 549  
 Db 3 CCCTTGCTCTGTGACTTCTCTGGGCTGCATCCATGTTCTCCGATCTGGCAGGG 62  
 QY 550 ArgAspGlnArgThrProGlnLysArgGluGluLeuValLeuArgValGlnGlyPro 569  
 Db 63 CCGGACAGCCAGCCCGCCAGAACCGCGGAGAGCTGTGCTCGGGTCCA-GGCCCG 121  
 QY 570 GluLeuIleSerLeuValGluLeuIleLeuAlaGluAlaGluThrArgSerGlnAspGly 589  
 Db 122 GAGCTCATCAGCTGTGTGAGCTGATCTGGCGGAGGCGAGACCGGAGCCAGACGG 181  
 QY 590 AspThrAlaAlaCysSerLeuIleGlnAlaArgLeuProLeuLeuSerCysCysCys 609  
 Db 182 GACACAGCGCGCTGACGCTCATCCAGGCGCGGCTGCCCTGCTCAGCTGCTGT 241  
 QY 610 GlyAspAspGluSerValArgLysValThrGluHisLeuSerGlyCysIleGlnGlnTrp 629  
 Db 242 GGGGACGATGAGAGTGTGAGAGAGTGACGAGACCTGTGAGCTGCATCCAGCAGTGG 301  
 QY 630 GlyAspSerValLeuGlyArgArgCysArgAspLeuLeuGlnLeuTrpLeuGlnArg 649  
 Db 302 GGAGACAGCTGTGTGGAGGGGCTGGCGAGACTTCTGTCTGACCTCTACTACAGCGG 361  
 QY 650 ProGluLeuArgValProValProgluValLeuLeuHisSerGluGlyAlaAlaSerSer 669  
 Db 362 CCGGAGCTGGCGGCTCCGCTGCTGAGGCTCCTACAGCAGAGGAGGCGCTGCCAGCAGC 421  
 QY 670 SerValCysLysLeuAspGlyLeuIleHisArgPheIleIleHisLeuAlaAspThrSer 689  
 Db 422 AGCCTCTCAAGCTGTGAGGAGCTCATCCACCGCTTCATCAGCCTCTTGGCGACACAGC 481  
 QY 690 AspSerArgAlaLeuGluAsnArgGlyAlaAspAlaSerMetAlaCysArgLysLeuAla 709  
 Db 482 GACTCCCGGCTGTGGAGAACCGAGGGGCGATGCCAGATGGCTCCCGGAGACTGGCG 541  
 QY 710 ValAlaHisProLeuLeuLeuLeuArgHisLeuProMetIleAlaAlaLeuLeuHisGly 729  
 Db 542 GGGGCGACCGCGCTGCTGCTGTCAGGACCGCCCATGATGGCGGCTCTGCACGGA 601  
 QY 730 ArgThrHisLeuAsnPheGlnGluPheArgGlnIleAsnHisLeuSerCysPheLeuHis 749  
 Db 602 CG-ACCCACCTCACTCCAGAGGTTCCGGCGAGAACCACTGAGCTGCTCTGCAC 660  
 QY 750 ValLeuGlyLeuLeuGluLeuLeuGlnProHisValPheArgSerGluHisGlnGlyAla 769  
 Db 661 GGTCTGGGCTGTGTGAGCTGTGTCAGCGGACGATGTCGAGGAGACCGAGGAGGCGG 720  
 QY 770 LeuTrpAspCysLeu-LeuSerPheIleArgLeuLeuLeuAsnArgArgLysSerSerArg 789  
 Db 721 CTGTGGAGCTGCTTGTGTGCTCTCATCCGCT-GTGCATGATTCAGGAAGTCTCTCCG 779  
 QY 789 gHis-LeuAlaAlaPheIleAsnLysPheValGlnPheIleHis 803  
 Db 780 ACATCTGTGCTGCTTCATCAACAAGTGTGCTGATTCATCAT 822

RESULT 11  
 LOCUS BG825357 800 bp mRNA linear EST 22-MAY-2001  
 DEFINITION 602747827F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4900615 5',  
 mRNA sequence.  
 ACCESSION BG825357  
 VERSION BG825357.1 GI:14172944  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 800)  
NIH-MGC <http://imgc.ncl.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLCM1794 row: f column: 08  
High quality sequence stop: 800.

FEATURES  
source

1. 800  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4900615"  
/clone\_lib="NIH\_MGC\_17"  
/tissue\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pORF7; Site\_1: EcoRI; Site\_2: XhoI; CDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 126 a 272 c 266 g 136 t  
ORIGIN

## Alignment Scores:

Pred. No.: 2,96e-118 Length: 800  
Score: 1278.50 Matches: 259  
Percent Similarity: 96.64% Conservative: 0  
Best Local Similarity: 96.64% Mismatches: 7  
Query Match: 24.39% Indels: 4  
DB: 12 Gaps: 1

US-09-929-769-7 (1-1029) x BG825357 (1-800)

QY 530 proSerAlaSerLeuAspPheLeuTrpAlaCysIleHisValProArgIleTrpGlnGly 549  
DB 3 CCTCTGCGCTCTGCGACTTCCTGGCGCTGCATCCATCTCTCGCATCTGGCAGGG 62  
QY 550 ArgAspGlnArgThrProGlnLysArgArgGlnGlnLeuValLeuArgValGlnGlyPro 569  
DB 63 CGGAGACCAAGGCAAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121  
QY 570 GlnLeuIleSerLeuValGlnLeuIleLeuAlaGlnLeuValLeuArgSerGlnAspGly 589  
DB 122 GAGCTATCAGCTCTGGAGAGCTGATCTCGCCGAGGAGAGAGAGAGAGAGAGAGAG 181  
QY 590 AspThrAlaAlaCysSerLeuIleGlnAlaArgLeuProLeuLeuLeuSerCysCys 609  
DB 182 GACACAGCCCGCTGACAGCTCATCCAGCGCCGCGCTGCTGCTGCTGCTGCTGCT 241  
QY 610 GlnAspAspGlnSerValAlaGlyValThrGlnHisLeuSerGlyCysIleGlnGlnTrp 629  
DB 242 GGGGACCATAGAGTGCAGGAAGGTGACGAGACACCTGTCAGAGCTGCATCCAGCACTGC 301  
QY 630 GlnAspSerValLeuGlnArgArgCysArgAspLeuLeuLeuGlnLeuTrpLeuGlnArg 649  
DB 302 GGAAGACAGCTGCTGGCAGAGCGCTGCGGAGACCTTCTCTGACGCTTACCTACAGCG 361  
QY 650 ProGlnLeuArgValProValProGlnValLeuLeuHisSerGlnGlnAlaAlaSerSer 669  
DB 362 CCGAGAGCTGGGGTCCCGCTGAGAGGCTCTATGACAGCGAAGGGGCTGCGAGAGC 421

QY 670 SerValCysLysLeuAspGlnLeuIleHisArgPheIleThrLeuLeuAlaAspThrSer 689  
DB 422 AGGCTGTGCAAGGTGGAGCGAGCTCATCCACCGCTTATCATCGCTCTGGGAGACACAG 481  
QY 690 AspSerArgAlaLeuGlnAsnArgGlyAlaAspAlaSerMetAlaCysArgGlyLeuAla 709  
DB 482 GACTCCCGGGCGTTGGAGAACCGAGGGGCGGATGCCAGCATGGCTTGGCCGGAAGCTGGCG 541  
QY 710 ValAlaHisProLeuLeuLeuLeuArgHisLeuProMetIleAlaLeuLeuHisGly 729  
DB 542 GTGGCGACCCCGCTGCTGCTGCTCAGGACCTCCCATGATCCGGCGCTCTGGCAGCG 601  
QY 730 ArgThrHisLeuAsnProGlnLeuArgGlnGlnAsnHisLeuSerCysPheLeuHis 749  
DB 602 CGCACCACCTCAACTCCAGAGATTCCGGCAGACAGAACACCTGAGCTCTCTGTCAGC 661  
QY 750 ValLeuGlnLeuLeuGlnLeuLeuGlnProHisValPheArgSerGlnHisGlnGly 769  
DB 662 GTGCTGGCGCTGTGGAGCTGCTGCAGCGCAGCTGTTCGAGCGAGCAGCAGGAGGCG 721  
QY 769 AlaTrpAspCysLeuLeuSerPheIleArgLeuLeuLeuAsnTrpArgLysSerSerArg 789  
DB 722 GCTGTGGAGTGGCTTCTGCTTCATCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
QY 789 GlnHisLeuAlaAlaPheIleAsn 796  
DB 781 ATCT---GGCTGCTCATCAAC 799

RESULT 12  
BG768219  
LOCUS  
DEFINITION  
602744272F1 NIH\_MGC\_49 Homo sapiens CDNA clone IMAGE:4877441 5',  
mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 858)  
NIH-MGC <http://imgc.ncl.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC/DCPD/DP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLCM1761 row: p column: 18  
High quality sequence stop: 750.

FEATURES  
source

1. 858  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4877441"  
/clone\_lib="NIH\_MGC\_49"  
/tissue\_type="melanotic melanoma, high MDR (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pORF7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library.!"



BASE COUNT	128	a	261	c	294	g	155	t
ORIGIN								
Alignment Scores:								
Pred. No.:	3.78e-118							Length: 858
Score:	1278.00							Matches: 273
Percent Similarity:	95.52%							Conservative: 4
Best Local Similarity:	94.14%							Mismatches: 6
Query Match:	24.38%							Indels: 0
DB:	12							Gaps: 0

```

|||||
Db 183 GAACACAGGTTGACCCCTGCTCTGAGCTTCTGCGCCATCCATGTTCTT 242
Oy 545 ArgIleTgInglYArGspGlnArGThrProGlnLysArGArgGluGluLeuValLeu 564
Db 243 CCGCATGTGGCAGGGGGGAGACACGACCCCGACAGAGCGGGAGAGAGCTGCTG 302
Oy 565 ArgValGInglYProGlnLeuLeuLeuSerLeuValGluLeuLeuValGluLeu 584
Db 303 CCGGCTCCAGGGCCCGAGCTCATCAGCTGTGTGAGCTGCTGCTGCGGAGAGAG 362
Oy 585 ArgSerGlnAspGlyAspThrAlaAlaCysSerLeuLeuLeuLeuLeuLeuLeu 604
Db 363 CGAGGACAGAGAGGGAGACACAGCGCTGCTGAGCTCATCCAGGCGCGCTGCTG 422
Oy 605 LeuSerCysCysCysGlyAspAspGlnSerValArgValValThrGluHisLeuSerGly 624
Db 423 CTCACGTGCTGCTGTGGGAGACATGAGTGTAGAGAGAGAGACACCTGTACAGC 482
Oy 625 CysIleGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArgAspLeuLeuGln 644
Db 483 TGCATCCACAGAGTGGGAGACAGCGTGTGGCAGCGCTGCGGAGACCTTCTCTG 542
Oy 645 LeuTrpLeuGlnArGProGlnLeuLeuArgValProValProGluValLeuHisSerGlu 664
Db 543 CTCCTACTACAGGCGGAGCGGAGCTGCGGCTGCGCTGAGTCTCTACTGACACAG 602
Oy 665 GlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuLeuHisArgPheLeu 684
Db 603 GGGGCTGCCAGAGCAGCGCTGTCAGAGCTGAGAGAGCTATCCACCGCTTCAACG 662
Oy 685 LeuAlaAspThrSerAspSerArgAlaLeu---GluAsnArgGlyAlaAspAlaSerMet 703
Db 663 CTTCGCGAGACACAGCCACCTCCCGGGCGTTGGCAGAACCGAGGGCGGATGCG 722
Oy 704 Ala-CysArgLysLeuAlaValAla-----HisProLeuLeuLeuArgHisLeu 720
Db 723 GCCCTCCCGGAAACTGGGGCGGGGGGCCCGCCCTGCTTCTGTGTCAGSACCTGG 782
Oy 721 -PrometIleAla 724
Db 783 CCCATGATCGCG 795

RESULT 14
LOCUS BG468191 728 bp mRNA linear EST 21-MAR-2001.
DEFINITION 602509766F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:464040 5',
ACCESSION BG468191
VERSION BG468191.1 GI:13400461
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 728)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CMI415 row: c column: 17
High quality sequence stop: 724.
Location/Qualifiers
1..728
/organism="Homo sapiens"
FEATURES
source

```

```

/db_xref="taxon:9606"
/clone="IMAGE:464040"
/clone_1ib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 122 a 244 c 114 t
ORIGIN
Alignment Scores:
pred. No.: 4 1e-112 Length: 728
Score: 1217.00 Matches: 238
Percent Similarity: 99.17% Conservative: 2
Best Local Similarity: 98.35% Mismatches: 2
Query Match: 23.22% Indels: 1
DB: 12 Gaps: 0
US-09-929-769-7 (1-1029) x BG468191 (1-728)
Oy 525 GluGlnArGpheaSPProSerAlaSerLeuAspPheLeuTrpAlaCysIleHisValPro 544
Db 2 GAACAGAGGTTGACCCCTGCTCTGAGCTTCTGCGCCATCCATGTTCTT 61
Oy 545 ArgIleTgInglYArGspGlnArGThrProGlnLysArGArgGluGluLeuValLeu 564
Db 62 CGCATGTGGCAGGGGGGAGACACGCGCACCCCGACAGAGCGGGAGAGAGCTGTG 121
Oy 565 ArgValGInglYProGlnLeuLeuLeuSerLeuValGluLeuLeuValGluLeu 584
Db 122 CCGGCTCCAGGGCCCGAGAGCTCATCAGCTGTGTGAGCTGATCTGTGGCGGAGAG 181
Oy 585 ArgSerGlnAspGlyAspThrAlaAlaCysSerLeuLeuLeuLeuLeuLeuLeu 604
Db 182 CGAGGACAGAGACCGGAGACACAGCGCTGAGCTCATCCAGGCGCGCTGCTG 241
Oy 605 LeuSerCysCysCysGlyAspAspGlnSerValArgValValThrGluHisLeuSerGly 624
Db 242 CTCACGTGCTGCTGTGGGAGACATGAGTGTAGAGAGAGTGCAGACCTGTACAGC 301
Oy 625 CysIleGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArgAspLeuLeuGln 644
Db 302 TGCATCCAGAGAGTGGGAGACAGCGTGTGGCAGCGCTGCGGAGACCTTCTCTG 361
Oy 645 LeuTrpLeuGlnArGProGlnLeuLeuArgValProValProGluValLeuHisSerGlu 664
Db 362 CTCCTACTACAGGCGGAGCGGAGCTGCGGCTGAGTCTCTACTGACACAG 421
Oy 665 GlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuLeuHisArgPheLeu 684
Db 422 GGGGCTGCCAGAGCAGCGCTGTCAGAGCTGAGAGACTATCCACCGCTTCAACG 481
Oy 685 LeuAlaAspThrSerAspSerArgAlaLeuGlyAsnArgGlyAlaAspAlaSerMetAla 704
Db 482 CTTCGCGAGACACAGCCACTCCCGGGCTGTGGAGAACCGAGGGCGGATGCGCACAT 541
Oy 705 CysArgLysLeuAlaValAlaHisProLeuLeuLeuLeuLeuLeuLeuLeu 724
Db 542 TGCAGGAGAGTGGCGGTGGGACCGCTGCTGCTGCTGAGCAGCTGCCAGATCGCG 601
Oy 725 AlaLeuLeuHisGlyArgThrHisLeuAsnArgPheGlnGluPheArgGlnGlnHis 744
Db 602 GCGCTCTGACGCGGC -AACCACTTCACCTTCAGAGAGTTCGGAGAGCAACACCTG 660
Oy 745 SerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuGlnProHisValPheArgSer 764
Db 661 AGCTGCTTCTGACAGTGTGCGCTGTGTGAGCTGTGACGCGGAGACGTATCCGACG 720

```

QY 765 GIuHIS 766  
 Db 721 GAGCAC 726  
 RESULT 15  
 BE792741  
 LOCUS 601585088F1 NIH\_MGC\_7 Homo sapiens CDNA clone IMAGE:393692 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE792741  
 VERSION BE792741.1 GI:10213939  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 899)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: rgs@biml.nih.gov  
 Tissue Procurement: DCTD/DTF  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/MLNL at: image.lnl.gov  
 Plate: LCMW88 row: 0 column: 21  
 High quality sequence stop: 771.  
 Location/Qualifiers  
 source 1..899  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="393692"  
 /clone\_id="NIH\_MGC\_7"  
 /tissue\_type="small cell carcinoma"  
 /cell\_line="MGC3"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pORF7; Site:1: XhoI; Site:2:  
 EcoRI; CDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5-  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-CDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 159 a 290 c 294 g 156 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7,79e-111 Length: 899  
 Score: 1206.00 Matches: 260  
 Percent Similarity: 90.69% Conservative: 3  
 Best Local Similarity: 89.66% Mismatches: 24  
 Query Match: 23.01% Indels: 8  
 DB: 12 Gaps: 1  
 US-09-929-769-7 (1-1029) x BE792741 (1-899)  
 QY 534 LeuAspPheLeuTrpAlaCysIleHisValProArgIleTrpGlnGlyArgAspGlnArg 553  
 Db 3 CTGAGACTTCCTCTGGGCTGCATCATGTCTCGCATCTGGACGAGCGGACCGCC 62  
 QY 554 ThrProGlnArgArgGlnGlnGlnValLeuArgValGlnGlnGlyProGlnLeuLeuSer 573  
 Db 63 ACCCGCGAGAGCGCGGAGAGCTGCTGCTGGGTCGA-GGCGCGAGCTCATTCAGC 121  
 QY 574 LeuValGlnLeuIleLeuAlaGlnAlaGlnThrArgSerGlnAspGlnAspThrAlaAla 593  
 Db 122 CTGGTGGAGCTGATCTCTGGCGGAGCGAGACGCGGAGCCAGGAGCGGAGCACACGCCGCC 181

QY 594 CysSerLeuIleGlnAlaArgLeuProLeuLeuSerCysCysGlyAspAspGln 613  
 Db 182 TGCAGCTTCATCCAGCGCCGCTGCCCTGCCTGCTGCTGCTGGGAGCATGAG 241  
 QY 614 SerValArgLysValThrGlnHisLeuSerGlyCysIleGlnGlnTrpGlyAspSerVal 633  
 Db 242 AGTTCAGGAAGGTACGAGAGACCTGTACAGCTTCATCCAGCATGGGAGAGAGCGCTG 301  
 QY 634 LeuGlyArgArgCysArgAspLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 653  
 Db 302 CTGGGAGAGCGCTGCCGAGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361  
 QY 654 ValProValProGlnValLeuLeuLeuHisSerGlnGlnGlnGlnGlnGlnGlnGlnGln 673  
 Db 362 GTGCCCGTGGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421  
 QY 674 LeuAspGlyLeuIleHisArgPheIleThrLeuLeuAlaAspThrSerAspSerArgAla 693  
 Db 422 CTGGAGGAGCTCATCCACCGCTTCATCCAGCTTCCTGGGAGACACCGAGCTCCCGGCG 481  
 QY 694 LeuGlnAsnArgGlyAlaAspAlaSerMetAlaCysArgLysLeuAlaValAlaHisPro 713  
 Db 482 TTGGAGAACCGAGGGCGGATGCCAGATGGCTCCGAGAACCTGGCGGTGGACCGCCG 541  
 QY 714 LeuLeuLeuLeuArgHisLeuProMetIleAlaAlaLeuLeuHisGlyArgThrHisLeu 733  
 Db 542 CTGCTGCTGCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598  
 QY 734 AsnPheGlnGluPheArgGlnGlnGlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeu 753  
 Db 599 AACTTCCAGAGGTTCCGCGAGACCACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658  
 QY 754 LeuGlnLeuLeuGlnProHisValPheArg-SerGlnHisGlnGlnGlnGlnGlnGlnGln 773  
 Db 659 CTGGAGCTGCTGAGCGCCGAGCTGTTCGCGAGGAGCAACGAGGAGGAGCTGTGGAGCTG 718  
 QY 773 SLeuLeuSerPheIleArgLeuLeuLeuAsnTrpArg-LysSerSerArgHisLeuAla 793  
 Db 719 CTT-CTGCTCTCATCCCGCTGCTGTG-AAATACAGGAGAACTCCCGCATCTGGGCTG 776  
 QY 793 LaphIleAsnLysPheValGlnPheIleHisLysTyrlIleThrTyrlAsnAlaProAla 813  
 Db 777 CTTTCATCAACAGTTTGTGCAAGTTCAATCCATTAAG---GTAAATATCCCAACAGGCA 833  
 QY 813 LalleSerPheLeuGlnLysHisAla.821  
 Db 834 GAGCATTTCTTCTTGAAGAGACGCG 859

Search completed: May 14, 2003, 02:29:53  
 Job time : 3359 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: May 7, 2003, 19:41:03 ; Search time 28 Seconds  
(without alignments)  
3381.945 Million cell updates/sec

Title: US-09-929-769-7

Perfect score: 5241  
Sequence: 1 MHILVHAMVILTLTGPPRA.....MDPSAQISEALRIHMEAYM 1029

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCCT\_NEW\_PUB pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5241	100.0	1029	US-10-033-245-22	Sequence 22, Appl
2	5241	100.0	1029	US-10-033-223-22	Sequence 22, Appl
3	5241	100.0	1029	US-10-033-167-22	Sequence 22, Appl
4	5241	100.0	1029	US-10-033-244-22	Sequence 22, Appl
5	5241	100.0	1029	US-10-033-435-22	Sequence 22, Appl
6	5241	100.0	1029	US-10-032-990-22	Sequence 22, Appl
7	5241	100.0	1029	US-09-929-769-7	Sequence 7, Appl
8	5241	100.0	1029	US-10-032-996-22	Sequence 22, Appl
9	5241	100.0	1029	US-10-033-396-22	Sequence 22, Appl
10	5241	100.0	1029	US-10-033-246-22	Sequence 22, Appl
11	5241	100.0	1029	US-10-033-301-22	Sequence 22, Appl
12	5241	100.0	1029	US-10-033-326-22	Sequence 22, Appl
13	2055	39.2	404	US-09-984-271-235	Sequence 235, App
14	1665	31.8	328	US-09-984-271-148	Sequence 148, App
15	530.5	10.1	161	US-09-764-891-4155	Sequence 4155, Ap
16	156	3.0	2472	US-09-815-242-5064	Sequence 5064, Ap
17	135	2.6	1987	US-10-133-382-6	Sequence 6, Appl
18	135	2.6	2013	US-10-133-382-2	Sequence 2, Appl
19	135	2.6	2014	US-10-132-382-8	Sequence 8, Appl

20	135	2.6	2040	9	US-10-132-382-4	Sequence 4, Appl
21	131.5	2.5	1257	9	US-10-109-324-2	Sequence 2, Appl
22	127.5	2.4	3353	10	US-09-888-615-64	Sequence 64, Appl
23	125	2.4	2383	9	US-10-082-830-260	Sequence 260, App
24	123	2.3	693	9	US-10-029-217A-4	Sequence 4, Appl
25	122.5	2.3	4999	9	US-09-976-059-14	Sequence 14, Appl
26	121	2.3	1315	9	US-09-990-046-10	Sequence 10, Appl
27	117.5	2.2	2045	9	US-09-736-968A-109	Sequence 109, App
28	117.5	2.2	2045	10	US-09-736-969A-95	Sequence 95, Appl
29	117.5	2.2	2045	10	US-09-736-960-92	Sequence 92, Appl
30	116.5	2.2	3798	9	US-10-014-717-6	Sequence 6, Appl
31	116	2.2	1147	9	US-10-259-864-1	Sequence 1, Appl
32	115	2.2	1012	10	US-09-876-527-16	Sequence 16, Appl
33	114	2.2	979	9	US-10-038-686-3	Sequence 3, Appl
34	114	2.2	1023	9	US-09-893-519A-14	Sequence 14, Appl
35	111.5	2.1	1698	9	US-10-080-943-4	Sequence 4, Appl
36	111	2.1	3056	8	US-08-984-090-2	Sequence 2, Appl
37	111	2.1	3056	9	US-10-307-077-1	Sequence 1, Appl
38	111	2.1	3056	9	US-10-059-585-41	Sequence 41, Appl
39	111	2.1	3056	9	US-10-175-225-2	Sequence 2, Appl
40	110.5	2.1	621	10	US-09-799-777-77	Sequence 77, Appl
41	110.5	2.1	3057	9	US-10-175-225-3	Sequence 3, Appl
42	108.5	2.1	633	10	US-09-815-242-12077	Sequence 12077, A
43	108.5	2.1	1189	9	US-09-738-626-4140	Sequence 4140, Ap
44	108	2.1	852	9	US-10-261-482-2	Sequence 2, Appl
45	108	2.1	852	9	US-10-282-837-14	Sequence 14, Appl

#### ALIGNMENTS

RESULT 1  
US-10-033-245-22  
Sequence 22, Application US/10033245  
Patent No. US20020160392A1  
GENERAL INFORMATION:  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P293081C7  
CURRENT APPLICATION NUMBER: US/10/033,245  
CURRENT FILING DATE: 2001-12-27  
PRIOR APPLICATION NUMBER: 60/095,325  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/112,851  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,145  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,511  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/115,558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,565  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,733  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119,341  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: 60/119,537  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/119,965

10/31/02

```

? PRIOR FILING DATE: 1999-02-12
? PRIOR APPLICATION NUMBER: 60/162,506
? PRIOR FILING DATE: 1999-10-29
? PRIOR APPLICATION NUMBER: 60/170,262
? PRIOR FILING DATE: 1999-12-09
? PRIOR APPLICATION NUMBER: 60/187,202
? PRIOR FILING DATE: 2000-03-03
? PRIOR APPLICATION NUMBER: PCT/US99/12252
? PRIOR FILING DATE: 1999-06-02
? PRIOR APPLICATION NUMBER: PCT/US99/28634
? PRIOR FILING DATE: 1999-12-01
? PRIOR APPLICATION NUMBER: PCT/US99/28551
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US00/03565
? PRIOR FILING DATE: 2000-02-11
? PRIOR APPLICATION NUMBER: PCT/US00/04414
? PRIOR FILING DATE: 2000-02-22
? PRIOR APPLICATION NUMBER: PCT/US00/05841
? PRIOR FILING DATE: 2000-03-02
? PRIOR APPLICATION NUMBER: PCT/US00/08439
? PRIOR FILING DATE: 2000-03-30
? PRIOR APPLICATION NUMBER: PCT/US00/14941
? PRIOR FILING DATE: 2000-05-30
? PRIOR APPLICATION NUMBER: PCT/US00/15264
? PRIOR FILING DATE: 2000-06-02
? PRIOR APPLICATION NUMBER: PCT/US00/32678
? PRIOR FILING DATE: 2000-12-01
? NUMBER OF SEQ ID NOS: 38
? SEQ ID NO 22
? LENGTH: 1029
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-033-245-22

```

```

Query Match      100.0%; Score 5241; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MHLVYHAMVILTLGPRADDEFOALDIWPEEKPTPTAFVDTSEALLPDKL 60
DB 1 MHLVYHAMVILTLGPRADDEFOALDIWPEEKPTPTAFVDTSEALLPDKL 60
QY 61 RMIRSEVRLVDALDLEPOQLLFVQSGIPVSSMSKLLQFDQVAHDPOLEONIM 120
DB 61 RMIRSEVRLVDALDLEPOQLLFVQSGIPVSSMSKLLQFDQVAHDPOLEONIM 120
QY 121 DKNYMAHLVEVQHERGASGGQTFHSLTASLPPRRDSTEAPKPSPEOPIGGRIKVT 180
DB 121 DKNYMAHLVEVQHERGASGGQTFHSLTASLPPRRDSTEAPKPSPEOPIGGRIKVT 180
QY 181 QLVLPREDLQAGMFLQIFLSPDPKRWSSPPRVALLQALQOELARVYQSPVPGI 240
DB 181 QLVLPREDLQAGMFLQIFLSPDPKRWSSPPRVALLQALQOELARVYQSPVPGI 240
QY 241 TVRVLAQALATLSSPHGALVMSHRSHPFLCPLRLQOLCOYORCPVDGESSLFLKVL 300
DB 241 TVRVLAQALATLSSPHGALVMSHRSHPFLCPLRLQOLCOYORCPVDGESSLFLKVL 300
QY 301 QMLQWLDSPEVEGGPLRAQRLMLASQASAGRRLSDVNGGLRLAEALAFRODLEVVSTV 360
DB 301 QMLQWLDSPEVEGGPLRAQRLMLASQASAGRRLSDVNGGLRLAEALAFRODLEVVSTV 360
QY 361 RAVIATILRSEQCSVEPDLISKVQLGLEVRSPHLEELTAFTFSATADAASPPACKPVV 420
DB 361 RAVIATILRSEQCSVEPDLISKVQLGLEVRSPHLEELTAFTFSATADAASPPACKPVV 420
QY 421 VVSSLLQEEPEPLAGKPGADGSLFAVRGLSPSSGLLVDMLEMDPEVSSCPDLQRL 480
DB 421 VVSSLLQEEPEPLAGKPGADGSLFAVRGLSPSSGLLVDMLEMDPEVSSCPDLQRL 480
QY 481 FSRKKGQAOVPSFRYLLTLFTTHOSSMPTLTHQCIKRVLLGKSGREQFDDASIDFLMAC 540
DB 481 FSRKKGQAOVPSFRYLLTLFTTHOSSMPTLTHQCIKRVLLGKSGREQFDDASIDFLMAC 540

```

```

QY 541 IHVRIWQGRDQRTPOKRRRELYLWVGPELISLVELLAEATRSODGDTAACSLIQAR 600
DB 541 IHVRIWQGRDQRTPOKRRRELYLWVGPELISLVELLAEATRSODGDTAACSLIQAR 600
QY 601 LPLLSCCGDDESVKRVYTHELSCGTOQKGDVSGRCRDLLOLYQORELRYPEVYL 660
DB 601 LPLLSCCGDDESVKRVYTHELSCGTOQKGDVSGRCRDLLOLYQORELRYPEVYL 660
QY 661 LHSEGAASSVCKLDGILHRTITLADTSDSRALENGADASMKRLAAVAPLRLRL 720
DB 661 LHSEGAASSVCKLDGILHRTITLADTSDSRALENGADASMKRLAAVAPLRLRL 720
QY 721 PMIAALLHGRTHLNFQEFROQNLSCFLHVLGLLELLQPHVFRSEHOGALWDCILSIFRL 780
DB 721 PMIAALLHGRTHLNFQEFROQNLSCFLHVLGLLELLQPHVFRSEHOGALWDCILSIFRL 780
QY 781 LINVKSSRHIAAFINKEVQFIHKYTYNPAATISFLQKHADPLHDSFQNSDLVMLKSL 840
DB 781 LINVKSSRHIAAFINKEVQFIHKYTYNPAATISFLQKHADPLHDSFQNSDLVMLKSL 840
QY 841 LAGLSLRDRLDRGDEEGEESAGSLPLVSVSLFTPLTAEMA PYMKRLSRGQTV 900
DB 841 LAGLSLRDRLDRGDEEGEESAGSLPLVSVSLFTPLTAEMA PYMKRLSRGQTV 900
QY 901 DLEVLSDIDEMSRRELLISFSTNLQRLMSSAECCRLAFSLARSMQNSSTIAAF 960
DB 901 DLEVLSDIDEMSRRELLISFSTNLQRLMSSAECCRLAFSLARSMQNSSTIAAF 960
QY 961 LPRFMYCLGSODPFVOTALRLNPEYALLCOEHAVALHRAFLVGMYGOMDPSQISEAL 1020
DB 961 LPRFMYCLGSODPFVOTALRLNPEYALLCOEHAVALHRAFLVGMYGOMDPSQISEAL 1020
QY 1021 RIHMEAVM 1029
DB 1021 RIHMEAVM 1029

```

```

RESULT 2
US-10-033-223-22
? Sequence 22, Application US/10033223
? Patent No. US20020164646A1
? GENERAL INFORMATION:
? APPLICANT: Botstein, David
? APPLICANT: Desnoyers, Luc
? APPLICANT: Ferrara, Napoleone
? APPLICANT: Fong, Sherman
? APPLICANT: Gao, Wei-Qiang
? APPLICANT: Goddard, Audrey
? APPLICANT: Gurney, Austin L.
? APPLICANT: Pan, James
? APPLICANT: Roy, Margaret Ann
? APPLICANT: Stewart, Timothy A.
? APPLICANT: Tumas, Daniel
? APPLICANT: Watanabe, Colin K.
? APPLICANT: Wood, William I.
? TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
? FILE REFERENCE: P2930R1C9
? CURRENT APPLICATION NUMBER: US/10/033,223
? PRIOR FILING DATE: 2001-12-27
? PRIOR APPLICATION NUMBER: 60/095,325
? PRIOR FILING DATE: 1998-08-04
? PRIOR APPLICATION NUMBER: 60/112,851
? PRIOR FILING DATE: 1998-12-16
? PRIOR APPLICATION NUMBER: 60/113,145
? PRIOR FILING DATE: 1998-12-16
? PRIOR APPLICATION NUMBER: 60/113,511
? PRIOR FILING DATE: 1998-12-22
? PRIOR APPLICATION NUMBER: 60/115,558
? PRIOR FILING DATE: 1999-01-12
? PRIOR APPLICATION NUMBER: 60/115,565
? PRIOR FILING DATE: 1999-01-12

```

```

: PRIOR APPLICATION NUMBER: 60/115,733
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/119,341
: PRIOR FILING DATE: 1999-02-09
: PRIOR APPLICATION NUMBER: 60/119,537
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: 60/119,965
: PRIOR FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: 60/162,506
: PRIOR FILING DATE: 1999-10-29
: PRIOR APPLICATION NUMBER: 60/170,262
: PRIOR FILING DATE: 1999-12-09
: PRIOR APPLICATION NUMBER: 60/187,202
: PRIOR FILING DATE: 2000-03-03
: PRIOR APPLICATION NUMBER: PCT/US99/12252
: PRIOR FILING DATE: 1999-06-02
: PRIOR APPLICATION NUMBER: PCT/US99/28634
: PRIOR FILING DATE: 1999-12-01
: PRIOR APPLICATION NUMBER: PCT/US99/28551
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US00/03565
: PRIOR FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: PCT/US00/05841
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: PCT/US00/08439
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: PCT/US00/14941
: PRIOR FILING DATE: 2000-05-30
: PRIOR APPLICATION NUMBER: PCT/US00/15264
: PRIOR FILING DATE: 2000-06-02
: PRIOR APPLICATION NUMBER: PCT/US00/32678
: PRIOR FILING DATE: 2000-12-01
: NUMBER OF SEQ ID NOS: 38
: SEQ ID NO 22
: LENGTH: 1029
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-033-223-22

```

```

Query Match      100.0%; Score 5241; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MHILVHAMVILITGPRADSEFOALDIWPEEKPLPAFVDTSEALLPDMKTL 60
DB 1 MHILVHAMVILITGPRADSEFOALDIWPEEKPLPAFVDTSEALLPDMKTL 60
QY 61 MINSEVRLVDAALQDLPEQOLLFFVQSGIPVSSMSKLIQFLDQAVAHDPOTLEONIM 120
DB 61 RMIRSEVRLVDAALQDLPEQOLLFFVQSGIPVSSMSKLIQFLDQAVAHDPOTLEONIM 120
QY 121 DKNTMAHLYEVOHERGASGGOTFHSLSLTPRRDSTEAPKPKSSPEOPIGGRIRVGT 180
DB 121 DKNTMAHLYEVOHERGASGGOTFHSLSLTPRRDSTEAPKPKSSPEOPIGGRIRVGT 180
QY 121 DKNTMAHLYEVOHERGASGGOTFHSLSLTPRRDSTEAPKPKSSPEOPIGGRIRVGT 180
DB 121 DKNTMAHLYEVOHERGASGGOTFHSLSLTPRRDSTEAPKPKSSPEOPIGGRIRVGT 180
QY 181 QLRVIGPREDDLAGMFLQIFPLSPDRMOSSSPRYALALQQLGGLARVVOGSEVPGCI 240
DB 181 QLRVIGPREDDLAGMFLQIFPLSPDRMOSSSPRYALALQQLGGLARVVOGSEVPGCI 240
QY 241 TVRVLOALATLTLSPHGALVMSHRSFHLACPLLRLQCOYRCVPODTGFSSFLKVL 300
DB 241 TVRVLOALATLTLSPHGALVMSHRSFHLACPLLRLQCOYRCVPODTGFSSFLKVL 300
QY 301 QMLQWLDSPGVGGPRLAQLRLMLASQASAGRRLSDVRGGLRLALBALAFRODLEVVSTV 360
DB 301 QMLQWLDSPGVGGPRLAQLRLMLASQASAGRRLSDVRGGLRLALBALAFRODLEVVSTV 360
QY 361 RAVATATLSSGOCSEVPLISKVLQGLIEVRSPHLEELLTAFFSTADAASFPAACKPV 420
DB 361 RAVATATLSSGOCSEVPLISKVLQGLIEVRSPHLEELLTAFFSTADAASFPAACKPV 420

```

```

QY 421 VVSSLLQEEBPLAGKPGADGSLAVALRGSSGLVDMLEMDPEVSSCPDQLRL 480
DB 421 VVSSLLQEEBPLAGKPGADGSLAVALRGSSGLVDMLEMDPEVSSCPDQLRL 480
QY 481 FSRKRGQAQVPSFRPYLLTFTHQSSWPTLHQCIRVLLGKSRQRPDSASLDPLMAC 540
DB 481 FSRKRGQAQVPSFRPYLLTFTHQSSWPTLHQCIRVLLGKSRQRPDSASLDPLMAC 540
QY 541 IHVPRINGRODRTQKRRRELVYLVQGPBELISVELLAEKETSOGSDTAACSLIAR 600
DB 541 IHVPRINGRODRTQKRRRELVYLVQGPBELISVELLAEKETSOGSDTAACSLIAR 600
QY 541 IHVPRINGRODRTQKRRRELVYLVQGPBELISVELLAEKETSOGSDTAACSLIAR 600
DB 541 IHVPRINGRODRTQKRRRELVYLVQGPBELISVELLAEKETSOGSDTAACSLIAR 600
QY 601 LPLLSGCCGDESVRYKTEHLSCIOQMGDSVLGRRCRDLLQYLQRPETRYVEVL 660
DB 601 LPLLSGCCGDESVRYKTEHLSCIOQMGDSVLGRRCRDLLQYLQRPETRYVEVL 660
QY 661 LHSEGAASSVCKIDGLIHRFTLLADTSDSRALENRGADASMACKRLAVAHPLLLRHL 720
DB 661 LHSEGAASSVCKIDGLIHRFTLLADTSDSRALENRGADASMACKRLAVAHPLLLRHL 720
QY 721 PMIAALGRTPLNROFERQONHLSCLFVLVGLLELQPHVFRSHOGALMDCILSFTRL 780
DB 721 PMIAALGRTPLNROFERQONHLSCLFVLVGLLELQPHVFRSHOGALMDCILSFTRL 780
QY 781 LLNRYKSSRHLLAFTNKVVOFTHKYTTYNAPAISFLQKHADPLDLDSFDSNLSVLMKSL 840
DB 781 LLNRYKSSRHLLAFTNKVVOFTHKYTTYNAPAISFLQKHADPLDLDSFDSNLSVLMKSL 840
QY 841 LAGLSLPSRDRDTRDGLDEGESESSAGSLPVSVSLFTPLTAAMAPYMKRLSGQYVE 900
DB 841 LAGLSLPSRDRDTRDGLDEGESESSAGSLPVSVSLFTPLTAAMAPYMKRLSGQYVE 900
QY 901 DLLEVLSDIDEMSRRRPILSFSTNLQRLMSAEECCRNLAFLSALSMQSPSIAAF 960
DB 901 DLLEVLSDIDEMSRRRPILSFSTNLQRLMSAEECCRNLAFLSALSMQSPSIAAF 960
QY 961 LPTFYMCISGODEFVYQTAALRLPEYALICQEAHVALHRAFLVGMGMDPSAQISEAL 1020
DB 961 LPTFYMCISGODEFVYQTAALRLPEYALICQEAHVALHRAFLVGMGMDPSAQISEAL 1020
QY 1021 RILHMEAVM 1029
DB 1021 RILHMEAVM 1029

```

# RESULT 3

```

US-10-033-167-22
: Sequence 22, Application US/10033167
: Publication No. US20020182618A1
: GENERAL INFORMATION:
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Ferrata, Napoleone
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Goddard, Audrey
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumes, Daniel
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: FILE REFERENCE: P2930R1C10
: CURRENT FILING DATE: 2001-12-27
: PRIOR APPLICATION NUMBER: 60/095,325
: PRIOR FILING DATE: 1998-08-04
: PRIOR APPLICATION NUMBER: 60/112,851
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,145

```





```

: CURRENT APPLICATION NUMBER: US/10/033,244
: CURRENT FILING DATE: 2001-12-27
: PRIOR APPLICATION NUMBER: 60/095,325
: PRIOR FILING DATE: 1998-08-04
: PRIOR APPLICATION NUMBER: 60/112,851
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,145
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,511
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: 60/115,558
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/115,565
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/115,733
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/119,341
: PRIOR FILING DATE: 1999-02-09
: PRIOR APPLICATION NUMBER: 60/119,537
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: 60/119,965
: PRIOR FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: 60/162,506
: PRIOR FILING DATE: 1999-10-29
: PRIOR APPLICATION NUMBER: 60/170,262
: PRIOR FILING DATE: 1999-12-09
: PRIOR APPLICATION NUMBER: 60/187,202
: PRIOR FILING DATE: 2000-03-03
: PRIOR APPLICATION NUMBER: PCT/US99/12252
: PRIOR FILING DATE: 1999-06-02
: PRIOR APPLICATION NUMBER: PCT/US99/28634
: PRIOR FILING DATE: 1999-12-01
: PRIOR APPLICATION NUMBER: PCT/US99/28551
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US00/03565
: PRIOR FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: PCT/US00/05841
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: PCT/US00/08439
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: PCT/US00/14941
: PRIOR FILING DATE: 2000-05-30
: PRIOR APPLICATION NUMBER: PCT/US00/15264
: PRIOR FILING DATE: 2000-06-02
: PRIOR APPLICATION NUMBER: PCT/US00/32678
: PRIOR FILING DATE: 2000-12-01
: NUMBER OF SEQ ID NOS: 38
: SEQ ID NO 22
: LENGTH: 1029
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-033-244-22

Query Match      100.0%; Score 5241; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      181 QLRVLGPEDDLACMFLOIPLSPDPRMOSSPPVALALQALGOELARVVGSSPEVPGI 240
QY      241 TVEVLQALATLLSSPHGALVMSHRSHPFLACPLRLQCOYORCVODTGFSSLEFLKVL 300
Db      241 TVEVLQALATLLSSPHGALVMSHRSHPFLACPLRLQCOYORCVODTGFSSLEFLKVL 300
QY      301 QMLQWLDSPGVEGGPRLAQLRLMLASQASAGRRLSDVNGGLRLAELAFAFQDLEVVSTV 360
Db      301 QMLQWLDSPGVEGGPRLAQLRLMLASQASAGRRLSDVNGGLRLAELAFAFQDLEVVSTV 360
QY      361 RAVIATLRSEGCQSVPEPDLISKVLQGLIEVRSHPLEELLTAFFSATAADAAPPACKPVY 420
Db      361 RAVIATLRSEGCQSVPEPDLISKVLQGLIEVRSHPLEELLTAFFSATAADAAPPACKPVY 420
QY      421 VVSSLLQEEEPPLAGCKPGADGSLAEAVRGPSSGLVMDLEMDPEVSSCCDLOLRLL 480
Db      421 VVSSLLQEEEPPLAGCKPGADGSLAEAVRGPSSGLVMDLEMDPEVSSCCDLOLRLL 480
QY      481 FSRKKGQAQVPSFRPYLLTFTTQSSWPTLHQCIKRVLLGKSGREDFPSASLDFLMAC 540
Db      481 FSRKKGQAQVPSFRPYLLTFTTQSSWPTLHQCIKRVLLGKSGREDFPSASLDFLMAC 540
QY      541 IHVPRIWQGDQRTPOKREBELVRVQGPPELLISVELILAEATRSODGDTAACSLIQAR 600
Db      541 IHVPRIWQGDQRTPOKREBELVRVQGPPELLISVELILAEATRSODGDTAACSLIQAR 600
QY      601 LPPLSCCGDDSVKRVTEHLSCIQQMGDSVLRGRCDLLQOLYLORELPVPEVL 660
Db      601 LPPLSCCGDDSVKRVTEHLSCIQQMGDSVLRGRCDLLQOLYLORELPVPEVL 660
QY      661 LHSEGAASSSVCKLIDGLIHRFITLLADTSDSRALENRGADASMACRLAVAHPLLLRL 720
Db      661 LHSEGAASSSVCKLIDGLIHRFITLLADTSDSRALENRGADASMACRLAVAHPLLLRL 720
QY      721 PMIAALLHGTHLNFQEFROONHLSGFLHYLGLELLQHPVFSSEHGALMDCLSLTRL 780
Db      721 PMIAALLHGTHLNFQEFROONHLSGFLHYLGLELLQHPVFSSEHGALMDCLSLTRL 780
QY      781 LNWYKRSRHLAAFINKEVQFIHKYIYNAPAAISFLQKHADPLHDISFQNSDLVMLKSL 840
Db      781 LNWYKRSRHLAAFINKEVQFIHKYIYNAPAAISFLQKHADPLHDISFQNSDLVMLKSL 840
QY      841 LAGLSLPSRDDRTDGLDEGEESAGSLPLVSVSLFTPLTAEMAAPYKRLSRGQTV 900
Db      841 LAGLSLPSRDDRTDGLDEGEESAGSLPLVSVSLFTPLTAEMAAPYKRLSRGQTV 900
QY      901 DLLEVSDIDEMSRRETELISFESTNLORLMSSAECCRNLAFLSRMONSPTAAAF 960
Db      901 DLLEVSDIDEMSRRETELISFESTNLORLMSSAECCRNLAFLSRMONSPTAAAF 960
QY      961 LPTFMVCLGSQDEFEVOTALRNLPEVALLCOEHAAYVLLHRAFLVGMYGMDPSAQLSEAL 1020
Db      961 LPTFMVCLGSQDEFEVOTALRNLPEVALLCOEHAAYVLLHRAFLVGMYGMDPSAQLSEAL 1020
QY      1021 RILHMEAYM 1029
Db      1021 RILHMEAYM 1029

RESULT 5
US-10-033-435-22
: Sequence 22, Application US/10033435
: Publication No. US20030027256A1
: GENERAL INFORMATION:
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Goddard, Audrey
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Roy, Margaret Ann

```

```

APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Matanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2930R1C5
CURRENT FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: US/10/033,435
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/112,851
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,145
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,511
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/115,558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119,341
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119,537
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/119,965
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: 60/162,506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170,262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/187,202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/14941
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
NUMBER OF SEQ ID NOS: 38
SEQ ID NO 22
LENGTH: 1029
TYPE: PRT
ORGANISM: Homo sapiens
US-10-033-435-22
Query Match 100.0%; Score 5241; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 121 DKNYMAHLVEYQHERGASGGOTFHSLLTASLPERRDSTEAPKPSSEOPDIPGCRIRVGT 180
DB 121 DKNYMAHLVEYQHERGASGGOTFHSLLTASLPERRDSTEAPKPSSEOPDIPGCRIRVGT 180
QY 181 QLRVLGPEBDLACMFLOIFPLSPDPRWSSPRVALALDQALQOELARVYOGSPYVPGI 240
DB 181 QLRVLGPEBDLACMFLOIFPLSPDPRWSSPRVALALDQALQOELARVYOGSPYVPGI 240
QY 241 TVRVLQALATLSSPHGALVMSHRSHPFLACPLRLQCOYQRCVPDDTGFSSFLKVL 300
DB 241 TVRVLQALATLSSPHGALVMSHRSHPFLACPLRLQCOYQRCVPDDTGFSSFLKVL 300
QY 301 QMLQWLDSPGVGGPLRAQRLMLASQASAGRRSLDVNGLLRLAEALAFRODLEEVSV 360
DB 301 QMLQWLDSPGVGGPLRAQRLMLASQASAGRRSLDVNGLLRLAEALAFRODLEEVSV 360
QY 361 RAVIATLRSGEQCEVEDLISKVLQGLIEVRSPLLEELTAFFSATADASPFPACKPVV 420
DB 361 RAVIATLRSGEQCEVEDLISKVLQGLIEVRSPLLEELTAFFSATADASPFPACKPVV 420
QY 421 VVSSLIQEEEPPLAGKPKADGGSLEAVRLGSSGLVDMLEMDPREVSSCPDLQRL 480
DB 421 VVSSLIQEEEPPLAGKPKADGGSLEAVRLGSSGLVDMLEMDPREVSSCPDLQRL 480
QY 481 FSRKRGQAQAVPSFRPYLLTFTHQSSWPTLHQCIRVLLGKSRQRFDPASLDFLWAC 540
DB 481 FSRKRGQAQAVPSFRPYLLTFTHQSSWPTLHQCIRVLLGKSRQRFDPASLDFLWAC 540
QY 541 IHVPRIMQGRQRTPOKRRRELVLRVQGPGLISIVELILAEAFRSDDGDTAACSLQAR 600
DB 541 IHVPRIMQGRQRTPOKRRRELVLRVQGPGLISIVELILAEAFRSDDGDTAACSLQAR 600
QY 601 LPLLLSCCCGDBESVRKYTEHLSGCIQWGDVYGRRCRDLLQDLYQRELRVPVEVL 660
DB 601 LPLLLSCCCGDBESVRKYTEHLSGCIQWGDVYGRRCRDLLQDLYQRELRVPVEVL 660
QY 661 LHSEGAASSVCKLDGLIHRFTLTLADTSDSRALENRGADASMACRLAANAHLRL 720
DB 661 LHSEGAASSVCKLDGLIHRFTLTLADTSDSRALENRGADASMACRLAANAHLRL 720
QY 721 PMTALLHGRTHLNFOEERROONHLSCLFHVGLLELLOPHVFRSEHOGALMDCILSFT 780
DB 721 PMTALLHGRTHLNFOEERROONHLSCLFHVGLLELLOPHVFRSEHOGALMDCILSFT 780
QY 781 LFNTRKSSRHIAFINKFVOFIHKYITYNAPAAISFLQKHADPLHDSFNSDLVMLKSL 840
DB 781 LFNTRKSSRHIAFINKFVOFIHKYITYNAPAAISFLQKHADPLHDSFNSDLVMLKSL 840
QY 841 LAGLSLPSRDRDTRGIDEGEESAGSLPVVSFLPTLTAEMAPYKRLSRGQTV 900
DB 841 LAGLSLPSRDRDTRGIDEGEESAGSLPVVSFLPTLTAEMAPYKRLSRGQTV 900
QY 901 DLEVLSDIDEMSRREPELISFTSTNQRMLSSAEECCRNIAFSLARSMONSSIAAF 960
DB 901 DLEVLSDIDEMSRREPELISFTSTNQRMLSSAEECCRNIAFSLARSMONSSIAAF 960
QY 961 LPTFMVCLGSODEFEVOTALRNLPYVALLCOEHAVALHRAFLVGMYGMDPSAQISEAL 1020
DB 961 LPTFMVCLGSODEFEVOTALRNLPYVALLCOEHAVALHRAFLVGMYGMDPSAQISEAL 1020
QY 1021 RILHMEAVM 1029
DB 1021 RILHMEAVM 1029

```

```

RESULT 6
US-10-032-990-22
; Sequence 22, Application US/10032990
; Publication No. US20030032060A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc

```

```

APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tamas, Daniel
APPLICANT: Watanabe, Collin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2930R1C11
CURRENT FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: US/10/032,990
PRIOR FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: 60/095,325
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/112,851
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,145
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,511
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/115,558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119,341
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119,537
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/119,965
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: 60/162,506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170,262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/187,202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/14941
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
NUMBER OF SEQ ID NOS: 38
SEQ ID NO 22
LENGTH: 1029
TYPE: PRT
ORGANISM: Homo sapiens
US-10-032-990-22
Query Match 100.0% Score 5241 DB 9: Length 1029:
Best Local Similarity 100.0% Pred. No. 0:
Matches 1029: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

```

QY 1 MHILVHNAVILTLGPPRADSEFOALLDIWPFPEKPLPTAFVYDTSEALLLDDWLK 60
DB 1 MHILVHNAVILTLGPPRADSEFOALLDIWPFPEKPLPTAFVYDTSEALLLDDWLK 60
QY 61 RMIRSEVRLVDAIODELPQQLLFVOSFGIPVSSMSKLLQFLDOAVAHDPQTEQNIM 120
DB 61 RMIRSEVRLVDAIODELPQQLLFVOSFGIPVSSMSKLLQFLDOAVAHDPQTEQNIM 120
QY 121 DKNYMAHIVEVQHEGASGQTFHSLTASLPFRDRSTAPKPKSSPEQITQGRIRVGT 180
DB 121 DKNYMAHIVEVQHEGASGQTFHSLTASLPFRDRSTAPKPKSSPEQITQGRIRVGT 180
QY 181 QLRVLGPEDDLAGMFLQIPIPLSPDPRWSSSPRYALQOALGQELARVQGSSEVPIC 240
DB 181 QLRVLGPEDDLAGMFLQIPIPLSPDPRWSSSPRYALQOALGQELARVQGSSEVPIC 240
QY 241 TVRVLOALATLLSSPHGALVMSMRSHFLACPLRLQLCQYORCVQPDGTGSSSLFLKVL 300
DB 241 TVRVLOALATLLSSPHGALVMSMRSHFLACPLRLQLCQYORCVQPDGTGSSSLFLKVL 300
QY 301 QMLQWLDSPGVGGPLRAQLRMLASQASGRRLSDRGGLRLAEALARQDLEVVSTV 360
DB 301 QMLQWLDSPGVGGPLRAQLRMLASQASGRRLSDRGGLRLAEALARQDLEVVSTV 360
QY 361 RAVIATLRSGECSVEPDLISKVLOGLIEVRSBHLELLTAFFSATADAAFPACKPVY 420
DB 361 RAVIATLRSGECSVEPDLISKVLOGLIEVRSBHLELLTAFFSATADAAFPACKPVY 420
QY 421 VVSSLLLOEEEPPLAGKPPADGGSLEAVRLGSSGLLVWLEMLDPEVVSQCDLOLRLL 480
DB 421 VVSSLLLOEEEPPLAGKPPADGGSLEAVRLGSSGLLVWLEMLDPEVVSQCDLOLRLL 480
QY 481 FSRKRGGAQVPSFRPYLLTFTHOSSMPTLHOCIRVLLGKSREORFPSSADLTMAC 540
DB 481 FSRKRGGAQVPSFRPYLLTFTHOSSMPTLHOCIRVLLGKSREORFPSSADLTMAC 540
QY 541 HVPRLMOGRDQRTPOKRREELVLRVQPELISVLELLAEATRSQDQDTAACSILQAR 600
DB 541 HVPRLMOGRDQRTPOKRREELVLRVQPELISVLELLAEATRSQDQDTAACSILQAR 600
QY 601 LPILLSCCCGDESVAKVTEHLSGCIQWGDVYLGRCDLLQLLYQRPRLVPPEVL 660
DB 601 LPILLSCCCGDESVAKVTEHLSGCIQWGDVYLGRCDLLQLLYQRPRLVPPEVL 660
QY 661 LHSSEGAASSVCKLDGLIHRFTLLADTSDSRLENGADASACKKLVVAPDLILRL 720
DB 661 LHSSEGAASSVCKLDGLIHRFTLLADTSDSRLENGADASACKKLVVAPDLILRL 720
QY 721 PMIAALLHGRTHLNFQEFROQNLSCFLHVLGILELLQHPVRSERHOGALMPCILSFIRL 780
DB 721 PMIAALLHGRTHLNFQEFROQNLSCFLHVLGILELLQHPVRSERHOGALMPCILSFIRL 780
QY 781 LLYNRKSSRHLLAFINKFVQFIHKYTTYVNAPAIISFLQKHADPLHDLSDNSDYLWKL 840
DB 781 LLYNRKSSRHLLAFINKFVQFIHKYTTYVNAPAIISFLQKHADPLHDLSDNSDYLWKL 840
QY 841 LAGLSLPSRDDPTDRGLDEGEESAGSLPVSVSLFTPLTAEMAAPYMKRLSRQCTVE 900
DB 841 LAGLSLPSRDDPTDRGLDEGEESAGSLPVSVSLFTPLTAEMAAPYMKRLSRQCTVE 900
QY 901 DLLEVLSDIDEMSRREPELISFEFTNLRUMSSAECCCNLAFLSLARSMQNSPSTAAAF 960
DB 901 DLLEVLSDIDEMSRREPELISFEFTNLRUMSSAECCCNLAFLSLARSMQNSPSTAAAF 960
QY 961 LPTFMYCIGSODFEVVQTLARNLPEYALLCOEHAVALHRAFLVGYGMQDPSAQISEAL 1020
DB 961 LPTFMYCIGSODFEVVQTLARNLPEYALLCOEHAVALHRAFLVGYGMQDPSAQISEAL 1020
QY 1021 RILHMEAVM 1029
DB 1021 RILHMEAVM 1029

```

RESULT 7  
US-09-929-769-7  
Sequence 7, Application US/09929769  
Publication No. US20030055224A1  
GENERAL INFORMATION:  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Polakis, Paul  
APPLICANT: Shou, Jianyong  
APPLICANT: Smith, Victoria  
APPLICANT: Soriano, Robert  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wu, Thomas D.  
APPLICANT: Zhang, Zhenh  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
FILE REFERENCE: P5007R1-US  
CURRENT APPLICATION NUMBER: US/09/929,769  
CURRENT FILING DATE: 2001-08-14  
PRIOR APPLICATION NUMBER: 60/089,653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/090,355  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/104,257  
PRIOR FILING DATE: 1998-10-14  
PRIOR APPLICATION NUMBER: 60/119,537  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/141,037  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: 60/162,506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: PCT/US99/12252  
PRIOR FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: PCT/US99/20111  
PRIOR FILING DATE: 1999-09-01  
PRIOR APPLICATION NUMBER: PCT/US99/28634  
PRIOR FILING DATE: 1999-12-01  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
PRIOR APPLICATION NUMBER: PCT/US00/00376  
PRIOR FILING DATE: 2000-01-06  
PRIOR APPLICATION NUMBER: PCT/US00/04342  
PRIOR FILING DATE: 2000-02-18  
PRIOR APPLICATION NUMBER: PCT/US00/08439  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: PCT/US00/13705  
PRIOR FILING DATE: 2000-05-17  
PRIOR APPLICATION NUMBER: PCT/US00/23328  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: PCT/US01/06520  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/20118  
PRIOR FILING DATE: 2001-06-22  
PRIOR APPLICATION NUMBER: 09/888,257  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 8  
SEQ ID NO 7  
LENGTH: 1029  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-929-769-7

Query Match 100.0% Score 5241: DB 9: Length 1029:  
Best Local Similarity 100.0% Pred. No. 0:  
Matches 1029: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 MHILVHANYILTLGPRADDSEFQALDIPPEEKPLPTAFVDTSEALLLPDKL 60  
DB 1 MHILVHANYILTLGPRADDSEFQALDIPPEEKPLPTAFVDTSEALLLPDKL 60

QY 61 RMIRSEVLRLVDAALDLEPOQLLEIVOSFCIPVSSMSKLLQFLDQAVAHDPOTLEONIM 120  
DB 61 RMIRSEVLRLVDAALDLEPOQLLEIVOSFCIPVSSMSKLLQFLDQAVAHDPOTLEONIM 120  
QY 121 DKNYMAHVEOHERGASGGOTFHSLLTASLPRRROSTEARPKPKSSPEOITGGCRIRVGT 180  
DB 121 DKNYMAHVEOHERGASGGOTFHSLLTASLPRRROSTEARPKPKSSPEOITGGCRIRVGT 180  
QY 181 QLRVLPGEEDDLACMFQILFPLSPDRWQSSPPRVALALQALGOELARVQSSPEVPGI 240  
DB 181 QLRVLPGEEDDLACMFQILFPLSPDRWQSSPPRVALALQALGOELARVQSSPEVPGI 240  
QY 241 TVRYLQALATLSSPHGALVMSMHSNHFACPLRLQCYQKVCPODTGFSSTFLKVL 300  
DB 241 TVRYLQALATLSSPHGALVMSMHSNHFACPLRLQCYQKVCPODTGFSSTFLKVL 300  
QY 301 QMLQWLDSPGVEGGPLRAQRLMLASQASAGRRLSDVAGGLRLAEALAFRODLEVSSTV 360  
DB 301 QMLQWLDSPGVEGGPLRAQRLMLASQASAGRRLSDVAGGLRLAEALAFRODLEVSSTV 360  
QY 361 RAVIATLRSGECCSVEPDLISKVQLIEVRSPLLELLTAFFSATADAAPPACKPVV 420  
DB 361 RAVIATLRSGECCSVEPDLISKVQLIEVRSPLLELLTAFFSATADAAPPACKPVV 420  
QY 421 VVSSLIQEEEPPLAGKPGADGSGLEAVRIGPSSGGLVMDLBMDEPVSVCEDLQRL 480  
DB 421 VVSSLIQEEEPPLAGKPGADGSGLEAVRIGPSSGGLVMDLBMDEPVSVCEDLQRL 480  
QY 481 FSRKKGQAQVPSFRYLLTLFTHOSSWPTLHOCIVLLGKSGREGFPDSASIDFLMAC 540  
DB 481 FSRKKGQAQVPSFRYLLTLFTHOSSWPTLHOCIVLLGKSGREGFPDSASIDFLMAC 540  
QY 541 IHVPRIMQGRDORTPOKREBELVLRVGPBELLISVELLAEATRSODGDTAACSLIQAR 600  
DB 541 IHVPRIMQGRDORTPOKREBELVLRVGPBELLISVELLAEATRSODGDTAACSLIQAR 600  
QY 541 IHVPRIMQGRDORTPOKREBELVLRVGPBELLISVELLAEATRSODGDTAACSLIQAR 600  
DB 541 IHVPRIMQGRDORTPOKREBELVLRVGPBELLISVELLAEATRSODGDTAACSLIQAR 600  
QY 601 LPLILSCCGDDSVKRVTEHLSCGTOQMGDSVYGRRCROLLQVLYQRELRVPVEVL 660  
DB 601 LPLILSCCGDDSVKRVTEHLSCGTOQMGDSVYGRRCROLLQVLYQRELRVPVEVL 660  
QY 661 LHSEGAASSVCKLDGLIHREFTLLADTSDSRALENRGADASMACKLAVAHPLLLRL 720  
DB 661 LHSEGAASSVCKLDGLIHREFTLLADTSDSRALENRGADASMACKLAVAHPLLLRL 720  
QY 721 PMIAALLHGRTHLNFQEFROQNLSCFLHVLGLLELLQPHVPNSEHOGALMDCILSIRL 780  
DB 721 PMIAALLHGRTHLNFQEFROQNLSCFLHVLGLLELLQPHVPNSEHOGALMDCILSIRL 780  
QY 781 LLNVRKSSRHAAFINKEVQFIHKYTYNAPPAISFLQKHADPLHDLSPNSDLVMLKSL 840  
DB 781 LLNVRKSSRHAAFINKEVQFIHKYTYNAPPAISFLQKHADPLHDLSPNSDLVMLKSL 840  
QY 841 LAGLSLPSRDDRDGDEEGEESASGLPIVSVSLFTPLTAEMAPYMKRSLRGOTVE 900  
DB 841 LAGLSLPSRDDRDGDEEGEESASGLPIVSVSLFTPLTAEMAPYMKRSLRGOTVE 900  
QY 901 DLEVLSDIDEMSRREPILSFSTNQLRLMSAECCRNLAFLSLARSQNSPSTAAAF 960  
DB 901 DLEVLSDIDEMSRREPILSFSTNQLRLMSAECCRNLAFLSLARSQNSPSTAAAF 960  
QY 961 LPTFMYCLGSODEFVQVOTALRNLPEVALCOEHAVALLHRAFLVGMGOMDPSAQISEAL 1020  
DB 961 LPTFMYCLGSODEFVQVOTALRNLPEVALCOEHAVALLHRAFLVGMGOMDPSAQISEAL 1020  
QY 1021 RILHMEAVM 1029  
DB 1021 RILHMEAVM 1029

RESULT 8  
US-10-032-996-22  
Sequence 22, Application US/10032996  
Publication No. US20030054447A1

## GENERAL INFORMATION:

APPLICANT: Botstein, David  
 APPLICANT: Desnoyers, Luc  
 APPLICANT: Ferrara, Napoleone  
 APPLICANT: Fong, Sherman  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Pan, James  
 APPLICANT: Roy, Margaret Ann  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumas, Daniel  
 APPLICANT: Watanabe, Colin K.  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 FILE REFERENCE: P293ORIC3  
 CURRENT FILING DATE: 2001-12-27  
 PRIOR APPLICATION NUMBER: 60/095,325  
 PRIOR FILING DATE: 1998-08-04  
 PRIOR APPLICATION NUMBER: 60/112,851  
 PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,145  
 PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,511  
 PRIOR FILING DATE: 1998-12-22  
 PRIOR APPLICATION NUMBER: 60/115,558  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,565  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,733  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/119,341  
 PRIOR FILING DATE: 1999-02-09  
 PRIOR APPLICATION NUMBER: 60/119,537  
 PRIOR FILING DATE: 1999-02-10  
 PRIOR APPLICATION NUMBER: 60/119,965  
 PRIOR FILING DATE: 1999-02-12  
 PRIOR APPLICATION NUMBER: 60/162,506  
 PRIOR FILING DATE: 1999-10-29  
 PRIOR APPLICATION NUMBER: 60/170,262  
 PRIOR FILING DATE: 1999-12-09  
 PRIOR APPLICATION NUMBER: 60/187,202  
 PRIOR FILING DATE: 2000-03-03  
 PRIOR APPLICATION NUMBER: PCT/US99/12252  
 PRIOR FILING DATE: 1999-06-02  
 PRIOR APPLICATION NUMBER: PCT/US99/28634  
 PRIOR FILING DATE: 1999-12-01  
 PRIOR APPLICATION NUMBER: PCT/US99/28551  
 PRIOR FILING DATE: 1999-12-02  
 PRIOR APPLICATION NUMBER: PCT/US00/03565  
 PRIOR FILING DATE: 2000-02-11  
 PRIOR APPLICATION NUMBER: PCT/US00/04414  
 PRIOR FILING DATE: 2000-02-22  
 PRIOR APPLICATION NUMBER: PCT/US00/05841  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: PCT/US00/08439  
 PRIOR FILING DATE: 2000-03-30  
 PRIOR APPLICATION NUMBER: PCT/US00/14941  
 PRIOR FILING DATE: 2000-05-30  
 PRIOR APPLICATION NUMBER: PCT/US00/15264  
 PRIOR FILING DATE: 2000-06-02  
 PRIOR APPLICATION NUMBER: PCT/US00/32678  
 PRIOR FILING DATE: 2000-12-01  
 NUMBER OF SEQ ID NOS: 38  
 SEQ ID NO 22  
 LENGTH: 1029  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-032-996-22

Query Match

100.0%; Score 5241; DB 9; Length 1029;

Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MHLIVHNAVILITLGGPRADSEFPALDIDIMPEEKPLPTAFIVNTSEALLIPWIKL	60
DB	1	MHLVHNAVILITLGGPRADSEFPALDIDIMPEEKPLPTAFIVNTSEALLIPWIKL	60
QY	61	RMIRSEVLRLVDAALDLEPPQULLFVQSGFIPVSSMSKLFLDQAVAHDPQLEONIM	120
DB	61	RMIRSEVLRLVDAALDLEPPQULLFVQSGFIPVSSMSKLFLDQAVAHDPQLEONIM	120
QY	121	DKNYMAHIVEVOHERGASGGQTFHSHLITLSPRRDSTAPPKSSPEQIGQGRIRVGT	180
DB	121	DKNYMAHIVEVOHERGASGGQTFHSHLITLSPRRDSTAPPKSSPEQIGQGRIRVGT	180
QY	181	QRLVIGPEDDLAGMFLQITFPLSPDPWQSSSRPVALLQALGQELARVVGSPVPGI	240
DB	181	QRLVIGPEDDLAGMFLQITFPLSPDPWQSSSRPVALLQALGQELARVVGSPVPGI	240
QY	241	TVRVIALATLLSPHGGALVMSHSHFLACPLRLQLCQYRCVQDTGFSFLFKVLL	300
DB	241	TVRVIALATLLSPHGGALVMSHSHFLACPLRLQLCQYRCVQDTGFSFLFKVLL	300
QY	301	OMLOWLDSPEVGGPLRAQLRLMLASQASAGRISDVRGGLRLAEALAERODLEVVSTV	360
DB	301	OMLOWLDSPEVGGPLRAQLRLMLASQASAGRISDVRGGLRLAEALAERODLEVVSTV	360
QY	361	RAVIATLRSGECCSVEPDLISKVLOGLIEVRSHLEELLTAFPSATADAAPPACKPVV	420
DB	361	RAVIATLRSGECCSVEPDLISKVLOGLIEVRSHLEELLTAFPSATADAAPPACKPVV	420
QY	421	VVSSLLQEEEPPLAGKPPADGSLAVRLGPSGLVWMLMDPEVSSCGDQLRL	480
DB	421	VVSSLLQEEEPPLAGKPPADGSLAVRLGPSGLVWMLMDPEVSSCGDQLRL	480
QY	481	FSRRKGGQAOVPSFRPYLLTFTHQSSWPTLHQICIRVLLGSRERFDPSSALDELMAC	540
DB	481	FSRRKGGQAOVPSFRPYLLTFTHQSSWPTLHQICIRVLLGSRERFDPSSALDELMAC	540
QY	541	IHYPRIMQGRDQRTPOKREELVLRVQGPGLISVLLIAEATRSQDDDTAACSILQAR	600
DB	541	IHYPRIMQGRDQRTPOKREELVLRVQGPGLISVLLIAEATRSQDDDTAACSILQAR	600
QY	601	LPLLISCCGGDESVAKVYEHLSGCIQWGDVSLGRCDLILQLYQRPRLRVPEVL	660
DB	601	LPLLISCCGGDESVAKVYEHLSGCIQWGDVSLGRCDLILQLYQRPRLRVPEVL	660
QY	661	LHSEGAASSVCKLQDGLIHRFTLLADTSDSRALENRGADASMACRKLAVAHPLLLRL	720
DB	661	LHSEGAASSVCKLQDGLIHRFTLLADTSDSRALENRGADASMACRKLAVAHPLLLRL	720
QY	721	PMIAALHGRTHLNPOEFQOHNLSCLFVHLGLLELLQHPHVESEHGALMDCLLSFTIL	780
DB	721	PMIAALHGRTHLNPOEFQOHNLSCLFVHLGLLELLQHPHVESEHGALMDCLLSFTIL	780
QY	781	LNTRYKSRHLLAFLINKFQFTHKYTYNAPALISFLQKHADPLHDLSDNSDLVLEKSL	840
DB	781	LNTRYKSRHLLAFLINKFQFTHKYTYNAPALISFLQKHADPLHDLSDNSDLVLEKSL	840
QY	841	LAGLSLPSRDDRTDRGLDEEGEESAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQVE	900
DB	841	LAGLSLPSRDDRTDRGLDEEGEESAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQVE	900
QY	901	DLLEVLSDIDEMSRRLPELISFSTNLQRLMSAECCCNLAFLSLRSMQNSPSTAAAF	960
DB	901	DLLEVLSDIDEMSRRLPELISFSTNLQRLMSAECCCNLAFLSLRSMQNSPSTAAAF	960
QY	961	LPTFMVCLGSDPEVYQVATLRNLPEYALLQEHAAVLLHRAFLVGYGQMDPSAQISEAL	1020
DB	961	LPTFMVCLGSDPEVYQVATLRNLPEYALLQEHAAVLLHRAFLVGYGQMDPSAQISEAL	1020
QY	1021	RIIHMEAVM 1029	
DB	1021	RIIHMEAVM 1029	

Db 1021 RILHMEAVM 1029

RESULT 9  
US-10-033-396-22

Sequence 22, Application US/10033396  
 Publication No. US2003007657A1  
 GENERAL INFORMATION:  
 APPLICANT: Botstein, David  
 APPLICANT: Desnoyers, David  
 APPLICANT: Ferrera, Napoleone  
 APPLICANT: Fong, Sherman  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Pan, James  
 APPLICANT: Roy, Margaret Ann  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumas, Daniel  
 APPLICANT: Watanabe, Collin K.  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 TITLE OF INVENTION: Acids Encoding the Same  
 FILE REFERENCE: P2930R1C4  
 CURRENT APPLICATION NUMBER: US/10/033,396  
 CURRENT FILING DATE: 2001-12-27  
 PRIOR APPLICATION NUMBER: 60/095,325  
 PRIOR FILING DATE: 1998-08-04  
 PRIOR APPLICATION NUMBER: 60/112,851  
 PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,145  
 PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,511  
 PRIOR FILING DATE: 1998-12-22  
 PRIOR APPLICATION NUMBER: 60/115,558  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,565  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,733  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/119,341  
 PRIOR FILING DATE: 1999-02-09  
 PRIOR APPLICATION NUMBER: 60/119,537  
 PRIOR FILING DATE: 1999-02-10  
 PRIOR APPLICATION NUMBER: 60/119,965  
 PRIOR FILING DATE: 1999-02-12  
 PRIOR APPLICATION NUMBER: 60/162,506  
 PRIOR FILING DATE: 1999-10-29  
 PRIOR APPLICATION NUMBER: 60/170,262  
 PRIOR FILING DATE: 1999-12-09  
 PRIOR APPLICATION NUMBER: 60/187,202  
 PRIOR FILING DATE: 2000-03-03  
 PRIOR APPLICATION NUMBER: PCT/US99/12252  
 PRIOR FILING DATE: 1999-06-02  
 PRIOR APPLICATION NUMBER: PCT/US99/28634  
 PRIOR FILING DATE: 1999-12-01  
 PRIOR APPLICATION NUMBER: PCT/US99/28551  
 PRIOR FILING DATE: 1999-12-02  
 PRIOR APPLICATION NUMBER: PCT/US00/03565  
 PRIOR FILING DATE: 2000-02-11  
 PRIOR APPLICATION NUMBER: PCT/US00/04414  
 PRIOR FILING DATE: 2000-02-22  
 PRIOR APPLICATION NUMBER: PCT/US00/05841  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: PCT/US00/08439  
 PRIOR FILING DATE: 2000-03-30  
 PRIOR APPLICATION NUMBER: PCT/US00/14941  
 PRIOR FILING DATE: 2000-05-30  
 PRIOR APPLICATION NUMBER: PCT/US00/15264  
 PRIOR FILING DATE: 2000-06-02  
 PRIOR APPLICATION NUMBER: PCT/US00/32678  
 PRIOR FILING DATE: 2000-12-01  
 NUMBER OF SEQ ID NOS: 38

SEQ ID NO 22  
 LENGTH: 1029  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-033-396-22

Query Match 100.0%; Score 5241; DB 9; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MHILVHNAVYLLTLPGRADDSFQALLDIMPPEKPLTAFLVDPSEALLPDLKTL 60  
 1 MHILVHNAVYLLTLPGRADDSFQALLDIMPPEKPLTAFLVDPSEALLPDLKTL 60  
 61 RMIRSEVLRVDAALDLEPQOLLFVQSGIPVSSSKLLQFLDQAVHDPQLEQNM 120  
 61 RMIRSEVLRVDAALDLEPQOLLFVQSGIPVSSSKLLQFLDQAVHDPQLEQNM 120  
 121 DKNYMAHLVEYQHERGASGGOTFHSLLTASLPFRDSTEARPKRSSPEQPIGGRIRVGT 180  
 121 DKNYMAHLVEYQHERGASGGOTFHSLLTASLPFRDSTEARPKRSSPEQPIGGRIRVGT 180  
 181 QLRVLGPEDDLQAGMFLQIFPLSPDRMSSPRVALALQALQGLARVQSGPEVPGI 240  
 181 QLRVLGPEDDLQAGMFLQIFPLSPDRMSSPRVALALQALQGLARVQSGPEVPGI 240  
 241 TVRVQLALATLLSSPHGALVMSMRSHFLACPLLRQLCOYQRCVPDQGFSSLEKVL 300  
 241 TVRVQLALATLLSSPHGALVMSMRSHFLACPLLRQLCOYQRCVPDQGFSSLEKVL 300  
 301 QMLQWLDSPVEGGPRLAQLMLASQASAGRRLSDVAGLLRLAFLAFLRDLLEVSVTV 360  
 301 QMLQWLDSPVEGGPRLAQLMLASQASAGRRLSDVAGLLRLAFLAFLRDLLEVSVTV 360  
 361 RAVATATRSRGECVPEPDLISKVLQGLIEVSPHLEELTAFFSATADAASPFPAKPVV 420  
 361 RAVATATRSRGECVPEPDLISKVLQGLIEVSPHLEELTAFFSATADAASPFPAKPVV 420  
 421 VVSSLLQEEEPRLAGKPGADGSLVAVRGSSGLLVDMLEMDPEVSSCPDLQRL 480  
 421 VVSSLLQEEEPRLAGKPGADGSLVAVRGSSGLLVDMLEMDPEVSSCPDLQRL 480  
 481 FSRKRGQAOVPSFRPYLLTLFTHOSSWPTLHOCIFVLGLKSKREORFDPASIDFLMAC 540  
 481 FSRKRGQAOVPSFRPYLLTLFTHOSSWPTLHOCIFVLGLKSKREORFDPASIDFLMAC 540  
 541 IHVRIWQGRDQRTPOKRREELVLRVQGPDLISVLELTAETRSODGDTAACSLIQAR 600  
 541 IHVRIWQGRDQRTPOKRREELVLRVQGPDLISVLELTAETRSODGDTAACSLIQAR 600  
 601 LPLLLSCCGDDESVKRVTEHLSCGCIQWGDVYGRRCRLDLLQYQREPLRVPEVYL 660  
 601 LPLLLSCCGDDESVKRVTEHLSCGCIQWGDVYGRRCRLDLLQYQREPLRVPEVYL 660  
 661 LHSEGAASSVCKLDGLIHRFTLLADTSDSRALENGADASMKRLAVALHPLLRHL 720  
 661 LHSEGAASSVCKLDGLIHRFTLLADTSDSRALENGADASMKRLAVALHPLLRHL 720  
 721 PMIAALLHGRTHLNFQFROQNLHSCFLHVLGLLELQPHVFRSEHOGALWDCLSFIRL 780  
 721 PMIAALLHGRTHLNFQFROQNLHSCFLHVLGLLELQPHVFRSEHOGALWDCLSFIRL 780  
 781 LLNRYKSSRHLLAFINKFVGFIRKITYYNAPAISFLQKHADPHDLSFSDSDVLMKSL 840  
 781 LLNRYKSSRHLLAFINKFVGFIRKITYYNAPAISFLQKHADPHDLSFSDSDVLMKSL 840  
 841 LAGLSLPSRDRTRGDEDEEESAGSLPVYSVLTPLTAEMAPYMKRLSRGQTV 900  
 841 LAGLSLPSRDRTRGDEDEEESAGSLPVYSVLTPLTAEMAPYMKRLSRGQTV 900  
 901 DLLEVLSDIDEMSRREPEILSFSTNLRQRLMSAECCRNLAFLSLRSMONSTIAA 960  
 901 DLLEVLSDIDEMSRREPEILSFSTNLRQRLMSAECCRNLAFLSLRSMONSTIAA 960

Qy 961 LPTFYCLSGDPEFVQVOTALRNLPEYALLCOEHAVALLRAPLVGYGMDSQAQISEAL 1020  
 |||||||  
 Db 961 LPTFYCLSGDPEFVQVOTALRNLPEYALLCOEHAVALLRAPLVGYGMDSQAQISEAL 1020

Qy 1021 RILHMEAVM 1029  
 |||||||  
 Db 1021 RILHMEAVM 1029

RESULT 10  
 US-10-033-246-22  
 ; Sequence 22, Application US/10033246  
 ; Patent No. US20020098505A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Botstein, David  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Ferrara, Napoleone  
 ; APPLICANT: Fong, Sherman  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Roy, Margaret Ann  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 ; TITLE OF INVENTION: Acids Encoding the Same  
 ; FILE REFERENCE: P2930R1C12  
 ; CURRENT FILING DATE: 2001-12-28  
 ; PRIOR APPLICATION NUMBER: US/10/033,246  
 ; PRIOR FILING DATE: 1998-08-04  
 ; PRIOR APPLICATION NUMBER: 60/095,325  
 ; PRIOR FILING DATE: 1998-08-04  
 ; PRIOR APPLICATION NUMBER: 60/112,851  
 ; PRIOR FILING DATE: 1998-12-16  
 ; PRIOR APPLICATION NUMBER: 60/113,145  
 ; PRIOR FILING DATE: 1998-12-16  
 ; PRIOR APPLICATION NUMBER: 60/113,511  
 ; PRIOR FILING DATE: 1998-12-22  
 ; PRIOR APPLICATION NUMBER: 60/115,558  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/115,565  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/115,733  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/119,341  
 ; PRIOR FILING DATE: 1999-02-09  
 ; PRIOR APPLICATION NUMBER: 60/119,537  
 ; PRIOR FILING DATE: 1999-02-10  
 ; PRIOR APPLICATION NUMBER: 60/119,965  
 ; PRIOR FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: 60/162,506  
 ; PRIOR FILING DATE: 1999-10-29  
 ; PRIOR APPLICATION NUMBER: 60/170,262  
 ; PRIOR FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: 60/187,202  
 ; PRIOR FILING DATE: 2000-03-03  
 ; PRIOR APPLICATION NUMBER: PCT/US99/12252  
 ; PRIOR FILING DATE: 1999-06-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28634  
 ; PRIOR FILING DATE: 1999-12-01  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28551  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US00/03565  
 ; PRIOR FILING DATE: 2000-02-11  
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414  
 ; PRIOR FILING DATE: 2000-02-22  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05841  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: PCT/US00/08439  
 ; PRIOR FILING DATE: 2000-03-30

;; PRIOR APPLICATION NUMBER: PCT/US00/14941  
 ;; PRIOR FILING DATE: 2000-05-30  
 ;; PRIOR APPLICATION NUMBER: PCT/US00/15264  
 ;; PRIOR FILING DATE: 2000-06-02  
 ;; PRIOR APPLICATION NUMBER: PCT/US00/32678  
 ;; PRIOR FILING DATE: 2000-12-01  
 ;; NUMBER OF SEQ ID NOS: 38  
 ;; SEQ ID NO 22  
 ;; LENGTH: 1029  
 ;; TYPE: PRT  
 ;; ORGANISM: Homo sapiens  
 US-10-033-246-22

Query Match 100.0%; score 5241; DB 12; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHLVHNAVILITLGPRAADSEFOALLDIWPEEKPPLPTAFVDTSEBALLPDMKL 60  
 |||||||  
 Db 1 MHLVHNAVILITLGPRAADSEFOALLDIWPEEKPPLPTAFVDTSEBALLPDMKL 60

Qy 61 RMIRSEVLRVDAALODLEPQQLLFVOSFGIPVSSMSKILQFLDQAVAHDPQTEONIM 120  
 |||||||  
 Db 61 RMIRSEVLRVDAALODLEPQQLLFVOSFGIPVSSMSKILQFLDQAVAHDPQTEONIM 120

Qy 121 DKNYMAHVEVOHERGASGQTFHSLTASLPRRDSTEAPKPKSSPEOPIGGRIRVGT 180  
 |||||||  
 Db 121 DKNYMAHVEVOHERGASGQTFHSLTASLPRRDSTEAPKPKSSPEOPIGGRIRVGT 180

Qy 181 QLRVLGPEDDLQAGMFLQIFPLSPDPWQSSPPPVALLQALQGLARVQGSPEVPGI 240  
 |||||||  
 Db 181 QLRVLGPEDDLQAGMFLQIFPLSPDPWQSSPPPVALLQALQGLARVQGSPEVPGI 240

Qy 241 TVRVQALATLTLSSPHGALVMSHRSHPFLACPLRQLQCYQRCVQDPTGFSLFLKVL 300  
 |||||||  
 Db 241 TVRVQALATLTLSSPHGALVMSHRSHPFLACPLRQLQCYQRCVQDPTGFSLFLKVL 300

Qy 301 QMLQWLDSPGVEGRLRAOLRMLASQASAGRRLSDYRGGLIRAEALAFRODLEVSVTV 360  
 |||||||  
 Db 301 QMLQWLDSPGVEGRLRAOLRMLASQASAGRRLSDYRGGLIRAEALAFRODLEVSVTV 360

Qy 361 RAVIATLRSGECCVPEPDLISKVLQGLIEYRSHBELTAFSSATADAASPPACKPVY 420  
 |||||||  
 Db 361 RAVIATLRSGECCVPEPDLISKVLQGLIEYRSHBELTAFSSATADAASPPACKPVY 420

Qy 421 VVSSLLOPEEPLAGGKPGADGSLFAVRIGPSSGLVWMLDELDPEVSSCPDLQRL 480  
 |||||||  
 Db 421 VVSSLLOPEEPLAGGKPGADGSLFAVRIGPSSGLVWMLDELDPEVSSCPDLQRL 480

Qy 481 FSRKRGQAQVPSFRPYLLTLETHOSWPTLHQCIRVLGKSREORFDPASLDFLMAC 540  
 |||||||  
 Db 481 FSRKRGQAQVPSFRPYLLTLETHOSWPTLHQCIRVLGKSREORFDPASLDFLMAC 540

Qy 541 IHVPRIMOGRDORTPKRREELVLRVQGPGLISLVELLAEATRSQDGTACSLQAR 600  
 |||||||  
 Db 541 IHVPRIMOGRDORTPKRREELVLRVQGPGLISLVELLAEATRSQDGTACSLQAR 600

Qy 601 IPLLSCCCGDDSVKRYTEHLSGCIQOMGDSVLGRCRDLILQTLQRPBELVAPPEVY 660  
 |||||||  
 Db 601 IPLLSCCCGDDSVKRYTEHLSGCIQOMGDSVLGRCRDLILQTLQRPBELVAPPEVY 660

Qy 661 LHSEGAASSVCKLDGLIHRFTLLADTSDRALENKRGADASNAACKLAVAPLILLRLH 720  
 |||||||  
 Db 661 LHSEGAASSVCKLDGLIHRFTLLADTSDRALENKRGADASNAACKLAVAPLILLRLH 720

Qy 721 PMTALLHGRTILNFOEFQONHLSCFLAVLGLLELLQPHVRSSEHGALMDCLSFIRL 780  
 |||||||  
 Db 721 PMTALLHGRTILNFOEFQONHLSCFLAVLGLLELLQPHVRSSEHGALMDCLSFIRL 780

Qy 781 LNTYKSSRHIAFTKPFQOFTHKTYTYPANAASIFLQKHADPLHLSPDNSDLVWLKSL 840  
 |||||||  
 Db 781 LNTYKSSRHIAFTKPFQOFTHKTYTYPANAASIFLQKHADPLHLSPDNSDLVWLKSL 840



QY 841 LAGLSLPSRDDTRDGLDEGESESSAGSLPLVSVSLFTPLTAENAPYKRLSGQTV 900  
DB 841 LAGLSLPSRDDTRDGLDEGESESSAGSLPLVSVSLFTPLTAENAPYKRLSGQTV 900  
QY 901 DLEEVSDIDEMSRRELLSFSTNLRMSAECCRLALSLRSMQNSPSTAAAF 960  
DB 901 DLEEVSDIDEMSRRELLSFSTNLRMSAECCRLALSLRSMQNSPSTAAAF 960  
QY 961 LPTFMYCLGSQDEEVYOTALRNLPEVALLCQEHAAVLHRAFLVGMVGMQDPSAQISEAL 1020  
DB 961 LPTFMYCLGSQDEEVYOTALRNLPEVALLCQEHAAVLHRAFLVGMVGMQDPSAQISEAL 1020  
QY 1021 RLHMEAVM 1029  
DB 1021 RLHMEAVM 1029

RESULT 11  
US-10-033-301-22  
Sequence 22, Application US/10033301  
Patent No. US20020098506A1  
GENERAL INFORMATION:  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
Acids Encoding the Same  
FILE REFERENCE: F2930R1C6  
CURRENT APPLICATION NUMBER: US/10/033,301  
CURRENT FILING DATE: 2001-12-27  
PRIOR APPLICATION NUMBER: 60/095,325  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/112,851  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,145  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,511  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/115,558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,565  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,733  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119,341  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: 60/119,537  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/119,965  
PRIOR FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: 60/162,506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 60/170,262  
PRIOR FILING DATE: 1999-12-09  
PRIOR APPLICATION NUMBER: 60/187,202  
PRIOR FILING DATE: 2000-03-03  
PRIOR APPLICATION NUMBER: PCT/US99/12252  
PRIOR FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: PCT/US99/28634  
PRIOR FILING DATE: 1999-12-01  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US00/03565

QY 1 PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
QY 2 PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: PCT/US00/05841  
QY 3 PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: PCT/US00/08439  
QY 4 PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: PCT/US00/14941  
QY 5 PRIOR FILING DATE: 2000-05-30  
PRIOR APPLICATION NUMBER: PCT/US00/15264  
QY 6 PRIOR FILING DATE: 2000-06-02  
PRIOR APPLICATION NUMBER: PCT/US00/32678  
QY 7 PRIOR FILING DATE: 2000-12-01  
NUMBER OF SEQ ID NOS: 38  
QY 8 SEQ ID NO 22  
LENGTH: 1029  
QY 9 TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-033-301-22

Query Match 100.0%; Score 5241; DB 12; Length 1029;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHILVHAMVITLTLGPRADDSFQALDIWPEEKPLPTAFVDTSEALLPMTKL 60  
DB 1 MHILVHAMVITLTLGPRADDSFQALDIWPEEKPLPTAFVDTSEALLPMTKL 60  
QY 61 RMIRSEVRLVDAADOLEPOQLLFVQSGFIVSSMSKLLQFLDQAVHPDPLEONIM 120  
DB 61 RMIRSEVRLVDAADOLEPOQLLFVQSGFIVSSMSKLLQFLDQAVHPDPLEONIM 120  
QY 121 DKNYMAHVEVQHERGASGGQTFHSLTASLPPRSDTEAPKPKSPEQPIGGRIKRG 180  
DB 121 DKNYMAHVEVQHERGASGGQTFHSLTASLPPRSDTEAPKPKSPEQPIGGRIKRG 180  
QY 181 QLRVLPEDDLACMFLOIFPLSPDPKQSSPPRVALALQALGQELARVQSSPEVPGI 240  
DB 181 QLRVLPEDDLACMFLOIFPLSPDPKQSSPPRVALALQALGQELARVQSSPEVPGI 240  
QY 241 TVRVLQALATLTLSSPHGALVMSMHRSHFLACPLLRQLCOYQKCPQDTGFSSILFKVLL 300  
DB 241 TVRVLQALATLTLSSPHGALVMSMHRSHFLACPLLRQLCOYQKCPQDTGFSSILFKVLL 300  
QY 301 QMLQWLDSPEVEGGPLRAQLRMLASQASAGRSLSDVKGILLRLAEALAFQDLEVVSTV 360  
DB 301 QMLQWLDSPEVEGGPLRAQLRMLASQASAGRSLSDVKGILLRLAEALAFQDLEVVSTV 360  
QY 361 RAVATATLRSGEQSVEDLISKVLOGLIEVRSPHLEELTAFFSATDAAAPPACKPVY 420  
DB 361 RAVATATLRSGEQSVEDLISKVLOGLIEVRSPHLEELTAFFSATDAAAPPACKPVY 420  
QY 421 VVSSILLQEEPEPLAGKPKGADGSLFAVRGSSGLVDMLEMDPVPVSSCPDLQRL 480  
DB 421 VVSSILLQEEPEPLAGKPKGADGSLFAVRGSSGLVDMLEMDPVPVSSCPDLQRL 480  
QY 481 FSRKKGQAQVPSFRPYLLTLTHOSSWPTLHOCIVLLGKSEQKQFDSASIDFLMAC 540  
DB 481 FSRKKGQAQVPSFRPYLLTLTHOSSWPTLHOCIVLLGKSEQKQFDSASIDFLMAC 540  
QY 541 IHVPRIMQGRDQRTPOKRRRELVLRVQGPETISVELLEAEATRSQDGTAAACSLQAR 600  
DB 541 IHVPRIMQGRDQRTPOKRRRELVLRVQGPETISVELLEAEATRSQDGTAAACSLQAR 600  
QY 601 LPLLSCCGDDSVKRVTEHLSCGICQOWGDSVLRRCRDLQLQVLRQELRVPEVL 660  
DB 601 LPLLSCCGDDSVKRVTEHLSCGICQOWGDSVLRRCRDLQLQVLRQELRVPEVL 660  
QY 661 LHSEGAASSVCKLDGLIHFTITLADTSSKRALENAGAASMAKRLAAVHPLLRHL 720  
DB 661 LHSEGAASSVCKLDGLIHFTITLADTSSKRALENAGAASMAKRLAAVHPLLRHL 720  
QY 721 PMIAALLHGRTHLNFQGFQROQNHLSCLFVHVLGLLELLQPVHVRSEHGALMDCILSIFRL 780



|||||  
Db 721 PMIAALHRTHLNFOEFGQNHLSCEFLHVLGELLELPHFVRSHEQGLMDCILSFIRL 780  
QY 781 LLYNRSSSHLAFLNKFFQFTHKYYTNAAPALSFLOKHADPLHLDSDNSLVMLKSL 840  
Db 781 LLYNRSSSHLAFLNKFFQFTHKYYTNAAPALSFLOKHADPLHLDSDNSLVMLKSL 840  
QY 841 LAGLSLPSHDDRTDGLDEGEESAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQYVE 900  
Db 841 LAGLSLPSHDDRTDGLDEGEESAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQYVE 900  
QY 901 DLLEVLSDIDEMSRRRPELISFEFTNLQRLMSAECCRNLAESLARMQNSPSIAAF 960  
Db 901 DLLEVLSDIDEMSRRRPELISFEFTNLQRLMSAECCRNLAESLARMQNSPSIAAF 960  
QY 961 LFTFWYCGSDPEFVYVOTLRNLPAYALLCOEHAVALLRAPLVGMYGMDPSAQISEAL 1020  
Db 961 LFTFWYCGSDPEFVYVOTLRNLPAYALLCOEHAVALLRAPLVGMYGMDPSAQISEAL 1020  
QY 1021 RILHMEAVM 1029  
Db 1021 RILHMEAVM 1029

RESULT 12  
US-10-033-326-22  
Sequence 22, Application US/10033326  
Patent No. US20020098507A1  
GENERAL INFORMATION:  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fond, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tamas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2930R1C8  
CURRENT APPLICATION NUMBER: US/10/033,326  
CURRENT FILING DATE: 2001-12-27  
PRIOR APPLICATION NUMBER: 60/095,325  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/112,851  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,145  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,511  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/115,558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,565  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,733  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119,341  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: 60/119,537  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/119,965  
PRIOR FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: 60/162,506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 60/170,262  
PRIOR FILING DATE: 1999-12-09  
PRIOR APPLICATION NUMBER: 60/187,202  
PRIOR FILING DATE: 2000-03-03

7/25/02

PRIOR APPLICATION NUMBER: PCT/US99/12252  
PRIOR FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: PCT/US99/28634  
PRIOR FILING DATE: 1999-12-01  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US00/03565  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: PCT/US00/05841  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: PCT/US00/08439  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: PCT/US00/14941  
PRIOR FILING DATE: 2000-05-30  
PRIOR APPLICATION NUMBER: PCT/US00/15264  
PRIOR FILING DATE: 2000-06-02  
PRIOR APPLICATION NUMBER: PCT/US00/32678  
PRIOR FILING DATE: 2000-12-01  
NUMBER OF SEQ ID NOS: 38  
SEQ ID NO 22  
LENGTH: 1029  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-033-326-22

Query Match 100.0%; Score 5241; DB 12; Length 1029;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLVYHAWYITLLTGPRAADSEFOALDIMPPEKPLPTAFVPTSEFALLPWLKL 60  
Db 1 MHLVYHAWYITLLTGPRAADSEFOALDIMPPEKPLPTAFVPTSEFALLPWLKL 60  
QY 61 RMIRSEVRLVDAALDLEFQOLLFVQSGFIPVSSNSKILQFLDQAVAHDPQLEQNM 120  
Db 61 RMIRSEVRLVDAALDLEFQOLLFVQSGFIPVSSNSKILQFLDQAVAHDPQLEQNM 120  
QY 121 DKNYMAHVEVQHERGASGGQFHSLLTASLPRRSTEARPKRSSPEQITGGRIYVGT 180  
Db 121 DKNYMAHVEVQHERGASGGQFHSLLTASLPRRSTEARPKRSSPEQITGGRIYVGT 180  
QY 181 QLRVLGPEDDLAQMFQIFPLSPDPRWQSSPPRVALALQALGOELARVYQSPPEVPGI 240  
Db 181 QLRVLGPEDDLAQMFQIFPLSPDPRWQSSPPRVALALQALGOELARVYQSPPEVPGI 240  
QY 241 TVRVQALATLLSSPHGALVMSHNSHTLACPLRLQLCYQRCVPODTGFSLSFLKYL 300  
Db 241 TVRVQALATLLSSPHGALVMSHNSHTLACPLRLQLCYQRCVPODTGFSLSFLKYL 300  
QY 301 QMLQWLDSPEVGGPLRAOLBMLASQASAGRRLSDVRGGLRLAEALAFQDLEVVSTV 360  
Db 301 QMLQWLDSPEVGGPLRAOLBMLASQASAGRRLSDVRGGLRLAEALAFQDLEVVSTV 360  
QY 361 RAVIATLRSGEQSVPEPDLISKVLQGLEVRSPHLEBELTAFPSADADAAPPACKPVY 420  
Db 361 RAVIATLRSGEQSVPEPDLISKVLQGLEVRSPHLEBELTAFPSADADAAPPACKPVY 420  
QY 421 VVSSLLQEEEPPLAGKPGADGSLAEAVRLGPSSGLVDMLEMLDEPVVSSCDLQRL 480  
Db 421 VVSSLLQEEEPPLAGKPGADGSLAEAVRLGPSSGLVDMLEMLDEPVVSSCDLQRL 480  
QY 481 FSRRRKGGAQVSPFRPYLLTFTTHOSSWPTLHOCIRVILIGKREORFQDSALDFLMAC 540  
Db 481 FSRRRKGGAQVSPFRPYLLTFTTHOSSWPTLHOCIRVILIGKREORFQDSALDFLMAC 540  
QY 541 IHVPRIWGRDQTPQKRREELVRYOGPELISLVELLIAEAFTRSODGDTAACSLIQAR 600  
Db 541 IHVPRIWGRDQTPQKRREELVRYOGPELISLVELLIAEAFTRSODGDTAACSLIQAR 600  
QY 601 LPLLSGCCGDDSVKRVTEHLSGCIQWGDVVGRRCDLLQLYLQRELPVPEVL 660  
Db 601 LPLLSGCCGDDSVKRVTEHLSGCIQWGDVVGRRCDLLQLYLQRELPVPEVL 660

Db 601 LPLLSCCGDDESVKRYTEHLSCGICQMGDSVIGRRCDLLLOLYORPELRVPVPEVL 660  
 Qy 661 LHSEGAASSVCKLDGLIHRFTITLADTSDSRALENRGADASMACRKLAVAHPLLLRL 720  
 Db 661 LHSEGAASSVCKLDGLIHRFTITLADTSDSRALENRGADASMACRKLAVAHPLLLRL 720  
 Qy 721 PMIAALLHGRTHLNFQEFROQNHLSCLFHVIGLLELLOPHVFRSEHOGALMDCLLSFIRL 780  
 Db 721 PMIAALLHGRTHLNFQEFROQNHLSCLFHVIGLLELLOPHVFRSEHOGALMDCLLSFIRL 780  
 Qy 781 LFNKRSRHLAAFINKFVQFIHKYITYNAPAAISFLQKHADPLHDLSFQNSDLYMKSL 840  
 Db 781 LFNKRSRHLAAFINKFVQFIHKYITYNAPAAISFLQKHADPLHDLSFQNSDLYMKSL 840  
 Qy 841 LAGLSRSDRTDRCGLDEEGEESAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQIVE 900  
 Db 841 LAGLSRSDRTDRCGLDEEGEESAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQIVE 900  
 Qy 901 DLEVLSDIDEMSRRPPEILSFSTNQLQRLMSAECCRNIAFSLALRSQNSPSIAAF 960  
 Db 901 DLEVLSDIDEMSRRPPEILSFSTNQLQRLMSAECCRNIAFSLALRSQNSPSIAAF 960  
 Qy 961 LPTMYCLGSODEFVYOTALRNLPYALCOEHAVALHRAFLVGMYGOMDPSAQISEAL 1020  
 Db 961 LPTMYCLGSODEFVYOTALRNLPYALCOEHAVALHRAFLVGMYGOMDPSAQISEAL 1020  
 Qy 1021 RLHMEAVM 1029  
 Db 1021 RLHMEAVM 1029

RESULT 13  
 US-09-984-271-235  
 : Sequence 235, Application US/09984271  
 : Publication No. US20030040088A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Rosen et al.  
 : TITLE OF INVENTION: 71 Human Secreted Proteins  
 : FILE REFERENCE: P2030P1  
 : CURRENT APPLICATION NUMBER: US/09/984, 271  
 : PRIOR FILING DATE: 2001-10-29  
 : PRIOR APPLICATION NUMBER: 09/482, 273  
 : PRIOR FILING DATE: 2000-01-13  
 : PRIOR APPLICATION NUMBER: PCT/US99/15849  
 : PRIOR FILING DATE: 1999-07-14  
 : PRIOR APPLICATION NUMBER: 60/092, 921  
 : PRIOR FILING DATE: 1998-07-15  
 : PRIOR APPLICATION NUMBER: 60/092, 922  
 : PRIOR FILING DATE: 1998-07-15  
 : PRIOR APPLICATION NUMBER: 60/092, 956  
 : PRIOR FILING DATE: 1998-07-15  
 : NUMBER OF SEQ ID NOS: 267  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 235  
 : LENGTH: 404  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 : US-09-984-271-235

Query Match 39.2%; Score 2055; DB 9; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-157;  
 Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 626 IQQMGDSVIGRRCDLLLOLYORPELRVPVPEVLHSEGAASSVCKLDGLIHRFTITL 685  
 Db 1 IQQMGDSVIGRRCDLLLOLYORPELRVPVPEVLHSEGAASSVCKLDGLIHRFTITL 60  
 Qy 686 ADTSDSRALENRGADASMACRKLAVAHPLLLRLPMIAALLHGRTHLNFQEFROQNHLS 745  
 Db 61 ADTSDSRALENRGADASMACRKLAVAHPLLLRLPMIAALLHGRTHLNFQEFROQNHLS 120  
 Qy 746 CFLHVLGELLELOPHVFRSEHOGALMDCLLSFIRLLNFKRSRHLAAFINKFVQFIHKY 805  
 Db 746 CFLHVLGELLELOPHVFRSEHOGALMDCLLSFIRLLNFKRSRHLAAFINKFVQFIHKY 805

Db 121 CFLHVLGELLELOPHVFRSEHOGALMDCLLSFIRLLNFKRSRHLAAFINKFVQFIHKY 180  
 Qy 806 ITYNAPAAISFLQKHADPLHDLSFQNSDLYMKSLLAGLSRSDRTDRCGLDEEGEES 865  
 Db 181 ITYNAPAAISFLQKHADPLHDLSFQNSDLYMKSLLAGLSRSDRTDRCGLDEEGEES 240  
 Qy 866 SAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQIVEDLEVLSDIDEMSRRPPEILSFST 925  
 Db 241 SAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQIVEDLEVLSDIDEMSRRPPEILSFST 300  
 Qy 926 NLQRLMSAECCRNIAFSLALRSQNSPSIAAFPLPFMYCCLGSODEFVYOTALRNLP 985  
 Db 301 NLQRLMSAECCRNIAFSLALRSQNSPSIAAFPLPFMYCCLGSODEFVYOTALRNLP 360  
 Qy 986 YALCOEHAVALHRAFLVGMYGOMDPSAQISEALRLHMEAVM 1029  
 Db 361 YALCOEHAVALHRAFLVGMYGOMDPSAQISEALRLHMEAVM 404

RESULT 14  
 US-09-984-271-148  
 : Sequence 148, Application US/09984271  
 : Publication No. US20030040088A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Rosen et al.  
 : TITLE OF INVENTION: 71 Human Secreted Proteins  
 : FILE REFERENCE: P2030P1  
 : CURRENT APPLICATION NUMBER: US/09/984, 271  
 : PRIOR FILING DATE: 2001-10-29  
 : PRIOR APPLICATION NUMBER: 09/482, 273  
 : PRIOR FILING DATE: 2000-01-13  
 : PRIOR APPLICATION NUMBER: PCT/US99/15849  
 : PRIOR FILING DATE: 1999-07-14  
 : PRIOR APPLICATION NUMBER: 60/092, 921  
 : PRIOR FILING DATE: 1998-07-15  
 : PRIOR APPLICATION NUMBER: 60/092, 922  
 : PRIOR FILING DATE: 1998-07-15  
 : PRIOR APPLICATION NUMBER: 60/092, 956  
 : PRIOR FILING DATE: 1998-07-15  
 : NUMBER OF SEQ ID NOS: 267  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 148  
 : LENGTH: 328  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 : FEATURE:  
 : NAME/KEY: SITE  
 : LOCATION: (328)  
 : OTHER INFORMATION: Xaa equals stop translation  
 : US-09-984-271-148

Query Match 31.8%; Score 1665; DB 9; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-126;  
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 703 MACRKLAVAHPLLLRLPMIAALLHGRTHLNFQEFROQNHLSCLFHVIGLLELOPHVF 762  
 Db 1 MACRKLAVAHPLLLRLPMIAALLHGRTHLNFQEFROQNHLSCLFHVIGLLELOPHVF 60  
 Qy 763 RSEHOGALMDCLLSFIRLLNFKRSRHLAAFINKFVQFIHKYITYNAPAAISFLQKHAD 822  
 Db 61 RSEHOGALMDCLLSFIRLLNFKRSRHLAAFINKFVQFIHKYITYNAPAAISFLQKHAD 120  
 Qy 823 PLHDLSPNSDLYMKSLLAGLSRSDRTDRCGLDEEGEESAGSLPLVSVSLFTPLT 882  
 Db 121 PLHDLSPNSDLYMKSLLAGLSRSDRTDRCGLDEEGEESAGSLPLVSVSLFTPLT 180  
 Qy 883 AAEMAPYMKRLSRGQIVEDLEVLSDIDEMSRRPPEILSFSTNLQRLMSAECCRNLA 942  
 Db 181 AAEMAPYMKRLSRGQIVEDLEVLSDIDEMSRRPPEILSFSTNLQRLMSAECCRNLA 240  
 Qy 943 FSLALRSQNSPSIAAFPLPFMYCCLGSODEFVYOTALRNLPYALCOEHAVALHRAFL 1002  
 Db 943 FSLALRSQNSPSIAAFPLPFMYCCLGSODEFVYOTALRNLPYALCOEHAVALHRAFL 1002

Db 241 ESLALRSMQNSPSIAAFLPTFMYCIGSODFEVQTALENLPEYALLCOEHAAVLLHRAF 300  
QY 1003 LVGMTGMDPSAQSISEALRIILHMEAVM 1029  
Db 301 LVGMTGMDPSAQSISEALRIILHMEAVM 327

## RESULT 15

US-09-764-891-4155  
; Sequence 4155, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIORITY FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4155  
; LENGTH: 161  
; TYPE: PRT  
; ORGANISM: Homo sapiens

FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (40)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids.  
; NAME/KEY: SITE  
; LOCATION: (84)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (94)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (96)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (98)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (105)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (136)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (137)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (139)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (141)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (143)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (145)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-4155

Query Match 10.1%; Score 530.5; DB 9; Length 161;

Best Local Similarity 73.1%; Pred. No. 5.5e-35; Indels 19; Caps 3;

Matches 117; Conservative 7; Mismatches 17;

QY 812 AAISFQKADPLADLSPDSDLVMLKSLAGLSLPSRDDRTDGLDEGEESAGSLP 871  
Db 2 AAISFQKADPLADLSPDSDLVMLKSLAGLSLPSRKDRTDGLDEGEESAGSLP 61  
QY 872 LVSVSLFTPLTAEMAPYMKRLSRGQTV-----DLLEVLSDIDEMSRREPEILS 921  
|||||

Db 62 LVSVSLFTPLTAEMAPYMKRLSRGQTVGESXPXSTPDLLLEXLSDIDEMSRREPEILS 121  
QY 922 FFSINLQRLMSAEE-----CCRNLAFLSLRSMQNSPSI 956  
Db 122 FFFDQ-----PAADELURPXKCPXLPALNSCRTPAL 157  
|||

Search completed: May 7, 2003, 19:43:42  
Job time : 33 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:37:58 ; Search time 44 Seconds

(Without alignments)  
3116.247 Million cell updates/sec

Title: US-09-929-769-7

Perfect score: 5241  
Sequence: 1 MHILVYHAMVILLTLCPPRA.....MDPSAQISEALRIHMEAVM 1029

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: A.Geneseq\_101002.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	5241	100.0	1029	21	Human PRO3434 prot
2	5241	100.0	1029	21	PRO3434, a novel s
3	5241	100.0	1029	23	AAU77765
4	3229	61.6	650	21	AAAB2495
5	2055	39.2	404	21	Human ORF2259
6	2055	39.2	404	22	Human ORF2259
7	2055	39.2	404	23	Human ORF2259
8	1665	31.8	327	22	Human gene 46 enco
9	1665	31.8	327	22	Human gene 46 enco
10	1665	31.8	327	22	Human gene 46 enco
					Human secreted pro
					Human secreted pro
					Human secreted pro

11	863	16.5	2042	22	AB859689
12	530.5	10.1	161	22	AAW95497
13	156	3.0	2472	22	AAU33568
14	149	2.8	1217	22	AAW52358
15	143	2.7	1519	21	AAW85660
16	140	2.7	1778	22	AAW94810
17	140	2.7	2099	22	AAW78496
18	138	2.6	3674	22	AAU14347
19	137.5	2.6	2757	22	ABG09921
20	135	2.6	2014	22	AAW67395
21	135	2.6	2014	22	AAW67395
22	133.5	2.5	931	22	AAW39815
23	133	2.5	737	15	AAW56494
24	133	2.5	737	17	AAW60804
25	133	2.5	737	18	AAW5019
26	132	2.5	2444	22	ABW62786
27	131	2.5	1302	22	AAW00761
28	128	2.4	1339	22	ABW21500
29	128	2.4	2473	22	AAW78911
30	127.5	2.4	3353	23	AAU82706
31	127	2.4	2733	22	ABW60737
32	127	2.4	3338	22	ABG09919
33	126	2.4	745	22	AAW41083
34	126	2.4	878	19	AAW54373
35	126	2.4	878	20	AAW53937
36	126	2.4	878	20	AAW53937
37	126	2.4	878	22	AAW53937
38	126	2.4	881	22	AAW31634
39	125.5	2.4	2779	22	ABW62371
40	125	2.4	901	22	AAU1540
41	125	2.4	2116	22	ABW58110
42	125	2.4	2383	23	ABW5631
43	125	2.4	2668	22	AAW79895
44	125	2.4	2797	22	AAU3159
45	125	2.4	3007	22	AAU28053

#### ALIGNMENTS

RESULT 1	AAW24027	standard; Protein; 1029 AA.
ID	AAW24027	
XX	AAW24027	
XX	25-JAN-2001	(first entry)
XX	Human PRO3434 protein sequence SEQ ID NO:23.	
XX	Human: tumour; diagnosis; neoplastic disease; proliferation; cancer;	
XX	Identification; tumorigenesis; anticancer; detection.	
XX	Homo sapiens.	
XX	MO200053750-A1.	
XX	14-SEP-2000.	
XX	02-DEC-1999;	99WO-US28551.
XX	08-MAR-1999;	99WO-US05028.
XX	01-SEP-1999;	99WO-US20111.
XX	29-OCT-1999;	99US-0162506.
XX	30-NOV-1999;	99WO-US28313.
XX	01-DEC-1999;	99WO-US28634.
XX	(GETH ) GENENTECH INC.	
XX	Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;	
XX	WPI: 2000-594320/56.	
XX	N-PSDB: AAC58109.	

XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit  
 PT the growth of tumors in mammals, and to identify inhibitors of PRO  
 PT polypeptide activity or expression -  
 PS  
 XX Claim 61; Fig 14; 226pp; English.  
 CC The present invention describes an antibody that binds to a human  
 CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO155; PRO1760;  
 CC PRO3434; PRO1293; PRO1295; PRO1303; PRO1344; PRO1354;  
 CC PRO3397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has  
 CC anticancer activity and can be used to diagnose tumors in mammals, by  
 CC detecting complex formation when the antibody is contacted with test  
 CC cells. Increased expression of genes encoding (I) can also be detected  
 CC to diagnose tumors. Agents which inhibit the activity of (I),  
 CC especially the antibodies, or an antisense oligonucleotide which  
 CC hybridises to genes encoding (I), can be used to inhibit tumor growth,  
 CC preferably by inducing cell death. Methods from the present invention  
 CC can be used to identify compounds which inhibit the biological activity  
 CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation  
 CC probes used in examples from the present invention for human PRO  
 CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human  
 CC PRO polynucleotide and protein sequences given in the exemplification of  
 CC the present invention.  
 CC  
 XX Sequence 1029 AA:  
 SO  
 Query Match 100.0%; Score 5241; DB:21; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 601 LPTLLSCCGDDSVKRVTEHLSGCIQWGDVSLGRRCRDLILQYLQPELVPPEVL 660  
 QY 661 LHSEGAASSVCKLDLIRHFTLLADTSDSRALENRGADASMACRKLAVAHPLLLRLH 720  
 DB 661 LHSEGAASSVCKLDLIRHFTLLADTSDSRALENRGADASMACRKLAVAHPLLLRLH 720  
 QY 721 PMTALLHGTHLNFOEFROQNLSCFLHYLGLLELLOPHVFSEHQALMDCLLSFIRL 780  
 DB 721 PMTALLHGTHLNFOEFROQNLSCFLHYLGLLELLOPHVFSEHQALMDCLLSFIRL 780  
 QY 781 LNVKRSRLAFIRKFOFIKITYNAPAFISFLQKHADPLHLSFQNSDLVWLKSL 840  
 DB 781 LNVKRSRLAFIRKFOFIKITYNAPAFISFLQKHADPLHLSFQNSDLVWLKSL 840  
 QY 841 LAGLSLPSRDRDRCGLDEGESESSAGSLPVSVLFTPLTAENAPYKRLRSQGYE 900  
 DB 841 LAGLSLPSRDRDRCGLDEGESESSAGSLPVSVLFTPLTAENAPYKRLRSQGYE 900  
 QY 901 DLEVLSDIDEMSRRPETLISFSTNLQRLMSAECCRNLAFLSLRSMQNSPTIAAF 960  
 DB 901 DLEVLSDIDEMSRRPETLISFSTNLQRLMSAECCRNLAFLSLRSMQNSPTIAAF 960  
 QY 961 LPTFMYCLSGODEEVYOTALRNLPEYALLCQEHAAVLHRAFLVGYMGOMDPSAQISEAL 1020  
 DB 961 LPTFMYCLSGODEEVYOTALRNLPEYALLCQEHAAVLHRAFLVGYMGOMDPSAQISEAL 1020  
 QY 1021 RILHMEAVM 1029  
 DB 1021 RILHMEAVM 1029

RESULT 2  
 AAY96736  
 ID AAY96736 standard; Protein; 1029 AA.  
 AC AAY96736;  
 XX  
 DT 26-SEP-2000 (first entry)  
 XX  
 DE PRO3434, a novel secreted protein.  
 XX  
 KW PRO3434; secreted protein; transmembrane protein; recombinant production;  
 KW gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Peptide  
 FT Location/Qualifiers  
 FT 1..16  
 FT /label= Signal\_peptide  
 FT 91..97  
 FT Modified-site  
 FT /note= "N-myristoylation site"  
 FT 136..142  
 FT Modified-site  
 FT /note= "N-myristoylation site"  
 FT 154..158  
 FT Modified-site  
 FT /note= "cAMP- and cGMP-dependent protein kinase  
 FT phosphorylation site"  
 FT 224..230  
 FT Modified-site  
 FT /note= "N-myristoylation site"  
 FT 329..333  
 FT Modified-site  
 FT /note= "amidation site"  
 FT 331..335  
 FT Modified-site  
 FT /note= "cAMP- and cGMP-dependent protein kinase  
 FT phosphorylation site"  
 FT 435..441  
 FT Modified-site  
 FT /note= "N-myristoylation site"  
 FT 439..445  
 FT Modified-site  
 FT /note= "N-myristoylation site"  
 FT 443..449  
 FT Modified-site  
 FT /note= "N-myristoylation site"  
 FT 616..620  
 FT Modified-site  
 FT /note= "cAMP- and cGMP-dependent protein kinase  
 FT phosphorylation site"  
 FT 634..636  
 FT Modified-site

```

FT      /note= "amidation site"
FT      665..671
FT      Modified-site
FT      /note= "N-myristoylation site"
FT      698..704
FT      Modified-site
FT      /note= "N-myristoylation site"
FT      785..789
FT      /note= "cAMP- and cGMP-dependent protein kinase
FT      phosphorylation site"
FT      891..895
FT      Modified-site
FT      /note= "cAMP- and cGMP-dependent protein kinase
FT      phosphorylation site"
FT      WO200036102-A2.
XX      22-JUN-2000.
XX      01-DEC-1999; 99WO-US28634.
XX      16-DEC-1998; 98US-0112851.
XX      16-DEC-1998; 98US-0113145.
XX      22-DEC-1998; 98US-0113511.
XX      12-JAN-1999; 99US-0115558.
XX      12-JAN-1999; 99US-0115565.
XX      12-JAN-1999; 99US-0115733.
XX      09-FEB-1999; 99US-0119341.
XX      10-FEB-1999; 99US-0119537.
XX      12-FEB-1999; 99US-0119965.
XX      02-JUN-1999; 99WO-US12252.
XX      (GETH ) GENENTECH INC.
XX      Botstein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A;
XX      Gueney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
XX      Wood WJ;
XX      WPI: 2000-431586/37.
XX      N-PSDB: AAA51266.
XX      Isolated nucleic acid molecule encodes a PRO polypeptide which is a
XX      transmembrane polypeptide
XX      Claim 12; Fig 16; 154pp: English.
XX      This is PRO3434, a novel secreted protein. The invention
XX      concerns novel secreted and transmembrane proteins, designated PRO.
XX      polypeptides. The cDNA and gene sequences are useful in the recombinant
XX      production of PRO polypeptides, as a hybridization probe to screen
XX      libraries to isolate cDNAs with sequence identity to PRO polypeptides or
XX      to map the gene encoding the PRO polypeptides and analyzing genetic
XX      disorders. The cDNA/gene can also be used to produce transgenic animals
XX      useful for the development and screening of therapeutically useful
XX      reagents. They can also be used in gene therapy, e.g. to replace a
XX      defective gene.
XX      Sequence 1029 AA:
SQ      Query Match 100.0%; Score 5241; DB 21; Length 1029;
        Best Local Similarity 100.0%; Pred. No. 0;
        Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB      181 QLRVLGPEDDLAGMFLQIFPLSPDPRMOSSSPRPVALALQALGQELARVVGSEPEVPGI 240
OY      241 TVRVLOALATLTLSPHGALVWMSHRSHFLACPLLROLCQYORCPOTGSSFLKVL 300
DB      241 TVRVLOALATLTLSPHGALVWMSHRSHFLACPLLROLCQYORCPOTGSSFLKVL 300
OY      301 QMLQWLDSPGVGGGLRLQRLMLASQASAGRSLSPVSGGLRLALALAFRODLEVVSTV 360
DB      301 QMLQWLDSPGVGGGLRLQRLMLASQASAGRSLSPVSGGLRLALALAFRODLEVVSTV 360
OY      361 RAVIATLRSGEQCSVEPDLISKVLQGLIEVRSPHLEELLTAFFSATDAASPACPKPV 420
DB      361 RAVIATLRSGEQCSVEPDLISKVLQGLIEVRSPHLEELLTAFFSATDAASPACPKPV 420
OY      421 VVSSLLQDEEPLAGKKGADGSGLEAVRLGSSGLVDMLMDPEVYSSCPDLQLRL 480
DB      421 VVSSLLQDEEPLAGKKGADGSGLEAVRLGSSGLVDMLMDPEVYSSCPDLQLRL 480
OY      481 FSRKKGQAOQVSPRPYLLTFTHQSSWPTLHQCIRVILGKSRQRPDPASLDPLMAC 540
DB      481 FSRKKGQAOQVSPRPYLLTFTHQSSWPTLHQCIRVILGKSRQRPDPASLDPLMAC 540
OY      541 IHVPRIMQGRDQRTDQKRREELVLRVQGPBELISVELLAEATRSQDGDPAACSLQAR 600
DB      541 IHVPRIMQGRDQRTDQKRREELVLRVQGPBELISVELLAEATRSQDGDPAACSLQAR 600
OY      601 LPLLSCCGDDESVRKYTEHLSGCIQMGDSVLRRCRDLLQLYORPELRVVPVYL 660
DB      601 LPLLSCCGDDESVRKYTEHLSGCIQMGDSVLRRCRDLLQLYORPELRVVPVYL 660
OY      661 LHSEGAASSVCKLKDGLHREFTLLADTSDSRALENRGADASMACRKLAVAHPLLLRL 720
DB      661 LHSEGAASSVCKLKDGLHREFTLLADTSDSRALENRGADASMACRKLAVAHPLLLRL 720
OY      721 PMIAALLHGRTHLNFQERROONHLSCLFVLGCLLELLOPHVRSRHOGALMDCLLSEFRL 780
DB      721 PMIAALLHGRTHLNFQERROONHLSCLFVLGCLLELLOPHVRSRHOGALMDCLLSEFRL 780
OY      781 LINTYKSSRHILAAFLINKFQFIHKYITYNAPAFISFLQKHADPLDHSFSDNSDLVMLSL 840
DB      781 LINTYKSSRHILAAFLINKFQFIHKYITYNAPAFISFLQKHADPLDHSFSDNSDLVMLSL 840
OY      841 LAGLSLPSRDRDTDGLDEEGEESASGLPLVSVSLFTPLTAAMAPYMKRLSGQTV 900
DB      841 LAGLSLPSRDRDTDGLDEEGEESASGLPLVSVSLFTPLTAAMAPYMKRLSGQTV 900
OY      901 DLLEVLSDIDEMSRRPETILSFSTNLQRLMSABECCRNLAFLSLALRSMONSPSIAAF 960
DB      901 DLLEVLSDIDEMSRRPETILSFSTNLQRLMSABECCRNLAFLSLALRSMONSPSIAAF 960
OY      961 LPTFWYCCGSDQFEVYQVATRLNLPETALICOHAHVLLHRAFLVGMYGMDPSAQISEAL 1020
DB      961 LPTFWYCCGSDQFEVYQVATRLNLPETALICOHAHVLLHRAFLVGMYGMDPSAQISEAL 1020
OY      1021 RLHMEAYN 1029
DB      1021 RLHMEAYN 1029

```

```

RESULT 3
ID      AAU77765
ID      AAU77765 standard; Protein: 1029 AA.
XX      AAU77765;
XX      05-JUN-2002 (first entry)
XX      Tumour associated antigenic target polypeptide (TAT) 156.
XX      Tumour associated antigenic target polypeptide: TAT; cancer:
XX      breast cancer; colorectal cancer; lung cancer; ovarian cancer;
XX      central nervous system cancer; liver cancer; bladder cancer;
KW

```





QY 781 L L N R K S S R H L A F I N K F V O F I H K Y I T Y N A P A I S F L O K H A D P L H D L S F D N S D L V M L K S L 840  
 DB 781 L L N R K S S R H L A F I N K F V O F I H K Y I T Y N A P A I S F L O K H A D P L H D L S F D N S D L V M L K S L 840  
 QY 841 L A G I S L S R D R T R G D E D E E E S S A G S L P L V S V S L F T P L T A A E M A P Y M K R L S R G Q T V E 900  
 DB 841 L A G I S L S R D R T R G D E D E E E S S A G S L P L V S V S L F T P L T A A E M A P Y M K R L S R G Q T V E 900  
 QY 901 D L E V L S D I D E M S R R P E I L S F F S T N I O R L M S S A E C C R N L A F S L A R S M O N S P S I A A F 960  
 DB 901 D L E V L S D I D E M S R R P E I L S F F S T N I O R L M S S A E C C R N L A F S L A R S M O N S P S I A A F 960  
 QY 961 L P T F M Y C L G S O D F E V V O T A L R N L P E Y A L L C Q E H A A V L L H R A F L V G M T G O M P S A Q I S E A L 1020  
 DB 961 L P T F M Y C L G S O D F E V V O T A L R N L P E Y A L L C Q E H A A V L L H R A F L V G M T G O M P S A Q I S E A L 1020  
 QY 1021 R I L H M E A V M 1029  
 DB 1021 R I L H M E A V M 1029

## RESULT 4

AAB42495  
 ID AAB42495 standard; Protein: 650 AA.

AC AAB42495;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2259 polypeptide sequence SEQ ID NO:4518.

XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vlneryary; antipsoptic; antiparkinsonian; noctropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX Homo sapiens.

PN WO200058473-A2.

PD 05-OCT-2000.

PE 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 9905-0127607.

PR 02-APR-1999; 9905-0127636.

PR 05-APR-1999; 9905-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI: 2000-602362/57.

DR N-PSDB; AAC76704.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS Claim 11; Page 3717-3718; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vlneryary;  
 CC antipsoptic; antiparkinsonian; noctropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiviral; antibacterial; antifungal; antirheumatic;  
 CC antihypertensive; and antineoplastic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SO Sequence 650 AA;

Query Match 61.6%; Score 3229; DB 21; Length 650;  
 Best Local Similarity 98.4%; Pred. No. 3.3e-289;  
 Matches 633; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

QY 397 E L L T A F S A T A D A S P F A C K P V V V S L L Q E E P L A G K P G A D G S L E A V R I G P S S G L 456  
 DB 8 E L L T A F S A T A D A S P F A C K P V V V S L L Q E E P L A G K P G A D G S L E A V R I G P S S G L 456  
 QY 457 L V D M L E M D P V V S S C P L Q L R L L F S R R K G K G A Q V P S F R Y L T L F H Q S W P T L H C I 516  
 DB 68 L V D M L E M D P V V S S C P L Q L R L L F S R R K G K G A Q V P S F R Y L T L F H Q S W P T L H C I 516  
 QY 517 R V L L G K S R E Q R P D P S A S I D F L M A C I H V R I W G R D Q R T P Q K R E E L V R V G P E L I S V E 576  
 DB 128 R V L L G K S R E Q R P D P S A S I D F L M A C I H V R I W G R D Q R T P Q K R E E L V R V G P E L I S V E 576  
 QY 577 L I L A E A F R S D G D T A A S L I Q A R L P L L S C C G D E S V R V T H L S C I O W K D S V L G R 636  
 DB 188 L I L A E A F R S D G D T A A S L I Q A R L P L L S C C G D E S V R V T H L S C I O W K D S V L G R 636  
 QY 637 R C R D L L D L Y L Q R P E L R V P E V L L H S E G A S S V Y C K I D G L I H R T T L A D T S D S R A L E N 696  
 DB 248 R C R D L L D L Y L Q R P E L R V P E V L L H S E G A S S V Y C K I D G L I H R T T L A D T S D S R A L E N 696  
 QY 697 R G A D A S M A C R K L A V A H P L L L R H L P M I A L L H G R T H L N F O E F R O O N H L S C F L A V L G L L E L 756  
 DB 308 R G A D A S M A C R K L A V A H P L L L R H L P M I A L L H G R T H L N F O E F R O O N H L S C F L A V L G L L E L 756  
 QY 757 L Q P H V F R S E H G A L M D C L S T F R L L N Y R K S S R H I A F I N K F V O F I H K Y I T Y N A P A I S F 816  
 DB 368 L Q P H V F R S E H G A L M D C L S T F R L L N Y R K S S R H I A F I N K F V O F I H K Y I T Y N A P A I S F 816  
 QY 817 L O K H A D P L H D L S F D N S D L V M L K S L A G I S L P S R D R T R G D E D E E E S S A G S L P L V S V S 876  
 DB 428 L O K H A D P L H D L S F D N S D L V M L K S L A G I S L P S R D R T R G D E D E E E S S A G S L P L V S V S 876  
 QY 877 L F T P L T A A E M A P Y M K R L S R G Q T V E - - - - - D L E V L S D I D E M S R R P E I L S F S T N 926  
 DB 488 L F T P L T A A E M A P Y M K R L S R G Q T V E S G P A S P T P D L L E V L S D I D E M S R R P E I L S F S T N 926  
 QY 927 L O R L M S S A E C C R N L A F S L A R S M O N S P S I A A F L P T F M Y C L G S O D F E V V O T A L R N L P E Y 986  
 DB 548 L O R L M S S A E C C R N L A F S L A R S M O N S P S I A A F L P T F M Y C L G S O D F E V V O T A L R N L P E Y 986  
 QY 987 A L L C O E H A A V L L H R A F L V G M T G O M P S A Q I S E A L R I L H M E A V M 1029  
 DB 608 A L L C O E H A A V L L H R A F L V G M T G O M P S A Q I S E A L R I L H M E A V M 1029

## RESULT 5

AA87196  
ID AAY87196 standard; Protein: 404 AA.  
AC AAY87196;  
XX  
XX 09-MAY-2000 (first entry)  
DE Human secreted protein sequence SEQ ID NO:235.  
XX  
XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
XX antinflammatory; nootropic; neuroprotective; anti-allergic; cancer;  
XX tumour; neurodegenerative disorder; developmental abnormality; allergy;  
XX foetal deficiency; blood disorder; immune system disorder; arthritis;  
XX autoimmune disease; hepatic disease; renal disease; inflammation;  
XX Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;  
XX infection; AIDS; spinal cord injury; transplant rejection; diabetes;  
XX asthma; sepsis; acne; psoriasis; cardiovascular disorder;  
XX reproductive disorder; gastrointestinal disorder; respiratory disorder;  
XX metabolic disorder; food additive; preservative.  
XX  
XX Homo sapiens.  
XX WO200004140-A1.  
XX  
XX 27-JAN-2000.  
XX  
XX 14-JUL-1999; 99WO-US15849.  
XX  
XX 15-JUL-1998; 98US-0092921.  
XX 15-JUL-1998; 98US-0092922.  
XX 15-JUL-1998; 98US-0092956.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;  
XX Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;  
XX Mucenski M, Endress GA, Soppet DR;  
XX WPI: 2000-161128/14.  
XX N-PSDB: AAY98062.  
XX  
XX New isolated human genes, useful for diagnosis and treatment of, e.g.  
XX cancers, neurological or blood disorders -  
XX  
XX Disclosure; Page 466-467; 494pp: English.  
XX  
XX The polynucleotide sequences given in AAY98017 to AAY98108 encode the  
XX human secreted proteins given in AAY87064 to AAY87223. Human secreted  
XX protein can have activities based on the tissues and cells the genes are  
XX expressed in. Examples of activities include: cytostatic;  
XX immunosuppressive; antinflammatory; nootropic; neuroprotective; and  
XX anti-allergic. The polynucleotides and their corresponding secreted  
XX polypeptides are useful for preventing, treating or ameliorating medical  
XX conditions, e.g. by protein or gene therapy. Also pathological conditions  
XX can be diagnosed by determining the amount of the new polypeptides in a  
XX sample or by determining the presence of mutations in the new  
XX polynucleotides. Human secreted proteins and their polynucleotides can  
XX be used for developing products for the diagnosis or treatment of cancer,  
XX tumours, neurodegenerative disorders, developmental abnormalities and  
XX foetal deficiencies, blood disorders, diseases of the immune system,  
XX autoimmune diseases, hepatic and renal disease, inflammation,  
XX allergies, Alzheimer's disease, behavioural disorders, schizophrenia,  
XX osteoporosis, arthritis, infections, AIDS, spinal cord injuries,  
XX transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,  
XX cardiovascular disorders, reproductive disorders, gastrointestinal  
XX disorders, respiratory disorders and metabolic disorders. The  
XX proteins or polynucleotides can also be used as food additives or  
XX preservatives. The proteins are also useful for identifying their  
XX binding partners. AAY98008 to AAY98016 and AAY87063 are sequence used in  
XX the exemplification of the present invention.  
XX  
XX Sequence 404 AA:

Query Match 39.2%; Score 2055; DB 21; Length 404;  
Best Local Similarity 100.0%; Pred. No. 6,1e-181;  
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 626 IQQWSDSVIGRRCDLLQYLRPELRVPPVLLHSGAASSVCKLDGLHRRITLL 665  
DB 1 IQQWSDSVIGRRCDLLQYLRPELRVPPVLLHSGAASSVCKLDGLHRRITLL 60  
QY 686 ADTSDSRALENRGADASMCRKLAVAHPLLLLRHLPMAALLGPHNLNFOEERQONHLS 745  
DB 61 ADTSDSRALENRGADASMCRKLAVAHPLLLLRHLPMAALLGPHNLNFOEERQONHLS 120  
QY 746 CPHLVGLLELLOPHYFRSEHOGALMDCLISFTRLNLNKRKSRHLAATNKVQITIKY 805  
DB 121 CPHLVGLLELLOPHYFRSEHOGALMDCLISFTRLNLNKRKSRHLAATNKVQITIKY 180  
QY 806 ITYNAPPAISFLQKHADPLHDSFSDSLVYMKSLAGSLSPRDRTPRGDEEGEES 865  
DB 181 ITYNAPPAISFLQKHADPLHDSFSDSLVYMKSLAGSLSPRDRTPRGDEEGEES 240  
QY 866 SAGSLPLVSVSLFTPLTAAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRRPETLSFFST 925  
DB 241 SAGSLPLVSVSLFTPLTAAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRRPETLSFFST 300  
QY 926 NLQRLMSAECCRNLAFLSLALRSQNSPSIAAFLPTMYCIGSDFEVYQALNLP 985  
DB 301 NLQRLMSAECCRNLAFLSLALRSQNSPSIAAFLPTMYCIGSDFEVYQALNLP 360  
QY 986 YALLCOEHAVALHRAFLVGMGOMPQSAQISALRLHMEAVN 1029  
DB 361 YALLCOEHAVALHRAFLVGMGOMPQSAQISALRLHMEAVN 404  
  
RESULT 6  
AAE06173  
ID AAE06173 standard; Protein: 404 AA.  
XX  
XX AAE06173;  
XX  
XX 24-SEP-2001 (first entry)  
XX  
XX Human gene 46 encoded secreted protein fragment, SpQ ID NO:235.  
XX  
XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;  
XX foetal abnormality; developmental abnormality; haematopoietic disorder;  
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
XX Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;  
XX psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;  
XX inflammation; neurological disorder; Alzheimer's disease; food additive;  
XX angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;  
XX pregnancy-related disorder; endocrine disorder; infection; wound healing;  
XX cell culture; chemotaxis; vulnery; binding partner identification;  
XX gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX WO200151504-A1.  
XX  
XX 19-JUL-2001.  
XX  
XX 12-JAN-2001; 2001WO-US00911.  
XX  
XX 13-JAN-2000; 2000US-0482273.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;  
XX Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;  
XX Endress GA, Mucenski M, Ebner R;  
XX WPI: 2001-425865/45.  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is

PT used in preventing, treating or ameliorating a medical condition  
XX  
PS Disclosure; Page 116; 864pp; English.

XX AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted  
CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.  
CC AAE06133-AAE06205 represent human secreted protein fragments.  
CC The secreted proteins and their genes are useful for preventing, treating  
CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
CC pathological conditions can be diagnosed by determining the amount of the  
CC new protein in a sample or by determining the presence of mutations in  
CC the new genes. Specific uses are described for each of the 71 genes,  
CC based on the tissues in which they are most highly expressed, and include  
CC developing products for the diagnosis or treatment of proliferative  
CC disorders, cancer, tumours, foetal and developmental abnormalities,  
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
CC pregnancy-related disorders, kidney disorders, gastrointestinal disorders,  
CC angiogenic disorders, endocrine disorders, and infections. The  
CC proteins can also be used to aid wound healing and epithelial cell  
CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
CC before transplantation, for supporting cell culture of primary tissues,  
CC to regenerate tissues, to identify their cognate ligands or binding  
CC partners, and in chemotaxis, and can be used as a food additive or  
CC preservative to modify storage properties. Antibodies specific for a  
CC protein of the invention can be used in alleviating symptoms associated  
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).  
CC The present sequence represents a human secreted protein fragment  
CC referred to in the disclosure of the invention.  
XX  
XX

Sequence 404 AA:

Query Match 39.2%; Score 2055; DB 22; Length 404;  
Best Local Similarity 100.0%; Pred. No. 6,1e-181;  
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 IQQMGSVIGRCRDLLDLYLQRPRLRVPEVLLHSGAASSVCKLDGLHRTTL 685  
DB 1 IQQMGSVIGRCRDLLDLYLQRPRLRVPEVLLHSGAASSVCKLDGLHRTTL 60  
QY 686 ADTSDBALENGGADASMAKRLAVAPLLLRHLPMLALHGRTHLNPOEERQNHLS 745  
DB 61 ADTSDBALENGGADASMAKRLAVAPLLLRHLPMLALHGRTHLNPOEERQNHLS 120  
QY 746 CFLHVLGELLELPYHFRSEHOGALMDCLSFIRLLNTRKSSRHIAAEINKFVQFIHRY 805  
DB 121 CFLHVLGELLELPYHFRSEHOGALMDCLSFIRLLNTRKSSRHIAAEINKFVQFIHRY 180  
QY 806 ITYNPAALISFIOKHAADPHLDLSFNSDLYMLKSLAGSLPSRDRTRGDEEBEES 865  
DB 181 ITYNPAALISFIOKHAADPHLDLSFNSDLYMLKSLAGSLPSRDRTRGDEEBEES 240  
QY 866 SAGSLPIVSVSFTPLTAEMAPYMKRLSRGQVEDLLFVLSPIDMSRRRPETISFEST 925  
DB 241 SAGSLPIVSVSFTPLTAEMAPYMKRLSRGQVEDLLFVLSPIDMSRRRPETISFEST 300  
QY 926 NIQRLMSSAECCRNLAESLARSQNSPSIAAFLPTMYCIGSODFEVOTALRNPDE 985  
DB 301 NIQRLMSSAECCRNLAESLARSQNSPSIAAFLPTMYCIGSODFEVOTALRNPDE 360  
QY 986 YALLCOEHAVALLHRAFLVGMGTGMDPSNOISEALRIIEMEAVM 1029  
DB 361 YALLCOEHAVALLHRAFLVGMGTGMDPSNOISEALRIIEMEAVM 404

RESULT 7  
ABG33997  
ID ABG33997 standard; Protein; 404 AA.  
XX

AC ABG33997;  
XX  
DT 15-JUL-2002 (first entry)  
XX

DE Human secreted protein #2 encoded by gene 46.

XX Human: secreted protein; gene therapy; immunosuppressive;  
XX antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiac;  
KW vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;  
KW virucide; fungicide; opthalmaological; autoimmune disease; neoplasm;  
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;  
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;  
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;  
KW ocular disorder; corneal infection; wound healing; skin aging;  
KW epithelial cell proliferation; food additive.

XX Homo sapiens.

OS WO200226931-A2.

PN 04-APR-2002.

PD 24-SEP-2001; 2001WO-US29871.

PF 25-SEP-2000; 2000US-234925P.

PR 12-JAN-2001; 2001WO-US00911.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;  
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;  
PI Edress GA, Mucenski M, Eder R;

XX WPI: 2002-362489/39.

DR N-PSDB: ABR69771.

XX Novel 71 isolated secreted polypeptides and polynucleotides encoding  
PT the polypeptides, useful for treating Huntington's disease, sepsis,  
PT meningitis, thrombocytopenia, haemolytic anaemia, rheumatoid arthritis,  
PT asthma  
XX

XX Disclosure; Page 1313-1314; 1478pp; English.

XX The invention relates to an isolated nucleic acid molecule (or its  
CC fragment, homologue complement or allelic variant) encoding a human  
CC secreted protein (and its fragment, domain, epitope, variant, secreted  
CC form and species variant). Also included are a recombinant vector  
CC comprising the nucleic acid, a recombinant host cell comprising the  
CC vector, an antibody against the secreted protein, a recombinant host cell  
CC that expresses the secreted protein and a method of identifying a binding  
CC partner of the secreted protein. The nucleic acid and protein are used to  
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.  
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep  
CC for example autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast, or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
CC ocular disorders e.g. corneal infection. Many other diseases and  
CC disorders are listed in the specification. The polypeptides can also be  
CC used to aid wound healing in epithelial cell proliferation, to prevent  
CC skin aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities. The present  
CC sequence represents a novel human secreted protein of the invention.  
XX

Sequence 404 AA:

Query Match 39.2%; Score 2055; DB 23; Length 404;  
Best Local Similarity 100.0%; Pred. No. 6,1e-181;  
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 IQQMGDSVLRRCRDLLOLYLQRPRLRVPEVYLHSEGAASSVCKDGLHRTFTLL 685  
 DB 1 IQQMGDSVLRRCRDLLOLYLQRPRLRVPEVYLHSEGAASSVCKDGLHRTFTLL 60  
 QY 686 ADSDSALENRGADAMACRKLAVAPHLRLPMTALLHGRTHLNQEPFQONHLS 745  
 DB 61 ADSDSALENRGADAMACRKLAVAPHLRLPMTALLHGRTHLNQEPFQONHLS 120  
 QY 746 CFLHVLGLLELLOPHVRSEHOGALMDCLSFIRLLNKRKSRHLLAFINKFVQFIHKY 805  
 DB 121 CFLHVLGLLELLOPHVRSEHOGALMDCLSFIRLLNKRKSRHLLAFINKFVQFIHKY 180  
 QY 806 ITYNAPAAISFLQKHAPRLDLSFNSDLYMKSLLAGLSLPSRSDRTDGLDEGEES 865  
 DB 181 ITYNAPAAISFLQKHAPRLDLSFNSDLYMKSLLAGLSLPSRSDRTDGLDEGEES 240  
 QY 866 SAGSLPVSVSLFTPLTAEMAPYMKRLSGQVYEDLEVLSDIDEMSRRPETLSFST 925  
 DB 241 SAGSLPVSVSLFTPLTAEMAPYMKRLSGQVYEDLEVLSDIDEMSRRPETLSFST 300  
 QY 926 NLQRLMSAECCRNLAFLSLARSMONSPSTIAAFLEPTFMYCIGSQDFEVVQALRNLP 985  
 DB 301 NLQRLMSAECCRNLAFLSLARSMONSPSTIAAFLEPTFMYCIGSQDFEVVQALRNLP 360  
 QY 986 YALLCOHAHVLRATLVNGMGMDPSAOISEALRTIHMEAVM 1029  
 DB 361 YALLCOHAHVLRATLVNGMGMDPSAOISEALRTIHMEAVM 404

RESULT 8  
 AAE06086  
 ID AAE06086 standard; Protein: 327 AA.  
 AC AAE06086;  
 XX  
 DT 24-SEP-2001 (first entry)  
 XX  
 DE Human gene 46 encoded secreted protein HUBC135, SEQ ID NO:148.  
 XX  
 KW Human: secreted protein; proliferative disorder; cancer; tumour; asthma;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;  
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;  
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;  
 KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;  
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;  
 KW cell culture; chemotaxis; vulnerrary; binding partner identification;  
 KW gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT /label= signal\_peptide  
 FT Protein 24..327  
 FT /label= Mature\_human\_secreted\_protein  
 XX  
 PN WO200151504-A1.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PE 12-JAN-2001; 2001WO-US00911.  
 XX  
 PR 13-JAN-2000; 2000US-0482273.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y,  
 PI Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;  
 PI Endress GA, Muscenski M, Ebner R;  
 XX  
 DR WPI: 2001-425865/45.

DR N-PSDB; AAD11675.  
 XX Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 PS Claim 11, page 771-772; 864pp; English.  
 XX  
 CC AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted  
 CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.  
 CC AAE06133-AAE06205 represent human secreted protein fragments.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 71 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).  
 CC The present sequence represents a human secreted protein of  
 CC the invention.  
 XX  
 SQ Sequence 327 AA;  
 XX  
 Query Match 31.8%; Score 1665; DB 22; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-145;  
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 703 MACRKLAVAPHLRLRLPMTALLHGRTHLNQEPFQONHLSCELVGLLELLOPHV 762  
 DB 1 MACRKLAVAPHLRLRLPMTALLHGRTHLNQEPFQONHLSCELVGLLELLOPHV 60  
 QY 763 RSEHOGALMDCLSFIRLLNKRKSRHLLAFINKFVQFIHKYITYNAPAAISFLQKHAD 822  
 DB 61 RSEHOGALMDCLSFIRLLNKRKSRHLLAFINKFVQFIHKYITYNAPAAISFLQKHAD 120  
 QY 823 PLHDLSDNSDLYMKSLLAGLSLPSRSDRTDGLDEGEESAGSLPVSVSLFTPLT 882  
 DB 121 PLHDLSDNSDLYMKSLLAGLSLPSRSDRTDGLDEGEESAGSLPVSVSLFTPLT 180  
 QY 883 AAEMAPYMKRLSGQVYEDLEVLSDIDEMSRRPETLSFSTNLQRLMSAECCRNLA 942  
 DB 181 AAEMAPYMKRLSGQVYEDLEVLSDIDEMSRRPETLSFSTNLQRLMSAECCRNLA 240  
 QY 943 FSLARSMONSPSTIAAFLEPTFMYCIGSQDFEVVQALRNLPYALLCOHAHVLRAT 1002  
 DB 241 FSLARSMONSPSTIAAFLEPTFMYCIGSQDFEVVQALRNLPYALLCOHAHVLRAT 300  
 QY 1003 LVNGMGMDPSAOISEALRTIHMEAVM 1029  
 DB 301 LVNGMGMDPSAOISEALRTIHMEAVM 327

RESULT 9  
 ABG33908  
 ID ABG33908 standard; Protein: 327 AA.  
 AC ABG33908;

XX 15-JUL-2002 (first entry)  
 XX Human secreted protein encode by gene 46.  
 DE  
 KW Human; secreted protein; gene therapy; immunosuppressive;  
 KW antithetic; antirheumatic; antiproliferative; cytostatic; cardiant;  
 KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;  
 KW virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;  
 KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;  
 KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;  
 KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;  
 KW ocular disorder; corneal infection; wound healing; skin aging;  
 KW epithelial cell proliferation; food additive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200226931-A2.  
 XX  
 PD 04-APR-2002.  
 XX  
 PF 24-SEP-2001; 2001WO-US29871.  
 XX  
 PR 25-SEP-2000; 2000US-234925P.  
 PR 12-JAN-2001; 2001WO-US00911.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;  
 PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;  
 PI Endress GA, Mucenski M, Ebner R;  
 XX  
 DR N-PSDB: ABR69771.  
 XX  
 PT Novel 71 isolated secreted polypeptides and polynucleotides encoding  
 PT the polypeptides, useful for treating Huntington's disease, sepsis,  
 PT meningitis, thrombocytopenia, haemolytic anaemia, rheumatoid arthritis,  
 PT asthma  
 XX  
 PS Claim 11: Page 1250-1251; 1478pp: English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (or its  
 CC fragment, homologue complement or allelic variant) encoding a human  
 CC secreted protein (and its fragment, domain, epitope, variant, secreted  
 CC form and species variant). Also included are a recombinant vector  
 CC comprising the nucleic acid, a recombinant host cell comprising the  
 CC vector, an antibody against the secreted protein, a recombinant host cell  
 CC that expresses the secreted protein and a method of identifying a binding  
 CC partner of the secreted protein. The nucleic acid and protein are used to  
 CC prevent, diagnose, treat or ameliorate a medical condition in e.g.  
 CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep  
 CC for example autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
 CC ocular disorders e.g. corneal infection. Many other diseases and  
 CC disorders are listed in the specification. The polypeptides can also be  
 CC used to aid wound healing an epithelial cell proliferation, to prevent  
 CC skin aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis. The polypeptides can also be used as a food additive or  
 CC preservative to increase or decrease storage capabilities. The present  
 CC sequence represents a novel human secreted protein of the invention.  
 XX  
 SQ Sequence 327 AA;  
 Query Match 31.8%; Score 1665; DB 23; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 5, 2e-145;  
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 703 MACRKLAVAHPLLLRLHPMTAALLHGRTHLNFGFEFRQONHLSCLFVHLGELLELPQHV 762

DB 1 MACRKLAVAHPLLLRLHPMTAALLHGRTHLNFGFEFRQONHLSCLFVHLGELLELPQHV 60  
 QY 763 RSEHOGALMDCLLSFIRLLNWRKSSRHLLAFINKFVFIKITYTNNPAAISFLQKAD 822  
 DB 61 RSEHOGALMDCLLSFIRLLNWRKSSRHLLAFINKFVFIKITYTNNPAAISFLQKAD 120  
 QY 823 PLHDLSPNSDLVMIKSLIAGLSLPSRDDTRDGLDEGEESAGSLPLVSVSLFPLT 882  
 DB 121 PLHDLSPNSDLVMIKSLIAGLSLPSRDDTRDGLDEGEESAGSLPLVSVSLFPLT 180  
 QY 883 AAEMAPYKRLRSQGVEDLLEVLSDIDEMSRREPEILSFESTNLQRLMSSAECCRNLA 942  
 DB 181 AAEMAPYKRLRSQGVEDLLEVLSDIDEMSRREPEILSFESTNLQRLMSSAECCRNLA 240  
 QY 943 FSLALRSMONSPSTAAALPTFMVCLGSDQDEEVQATARNLPEVALLCOEHAVALHRAAF 1002  
 DB 241 FSLALRSMONSPSTAAALPTFMVCLGSDQDEEVQATARNLPEVALLCOEHAVALHRAAF 300  
 QY 1003 LVGMVGMPSAQISEALRIHMEAVM 1029  
 DB 301 LVGMVGMPSAQISEALRIHMEAVM 327  
 RESULT 10  
 AAY87109  
 ID AAY87109 standard; Protein: 328 AA.  
 AC  
 AC AAY87109;  
 DT 09-MAY-2000 (first entry)  
 DT  
 DT  
 XX  
 DE Human secreted protein sequence SEQ ID NO:148.  
 DE  
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
 KW antiinflammatory; nootropic; neuroprotective; antiallergic; cancer;  
 KW tumour; neurodegenerative disorder; developmental abnormality; allergy;  
 KW focal deficiency; blood disorder; immune system disorder; arthritis;  
 KW autoimmune disease; hepatic disease; renal disease; inflammation;  
 KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;  
 KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;  
 KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;  
 KW reproductive disorder; gastrointestinal disorder; respiratory disorder;  
 KW metabolic disorder; food additive; preservative.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200004140-A1.  
 XX  
 PD 27-JAN-2000.  
 XX  
 PF 14-JUL-1999; 99WO-US15849.  
 XX  
 PR 15-JUL-1998; 98US-0092921.  
 PR 15-JUL-1998; 98US-0092922.  
 PR 15-JUL-1998; 98US-0092956.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;  
 PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;  
 PI Mucenski M, Endress GA, Soppet DR;  
 XX  
 DR WPI: 2000-161128/14.  
 DR N-PSDB: AA296062.  
 XX  
 PT New isolated human genes, useful for diagnosis and treatment of, e.g.  
 PT cancers, neurological or blood disorders  
 PT  
 PS Claim 11: Page 404-406; 494pp: English.  
 XX  
 CC The polynucleotide sequences given in AAY96017 to AAY98108 encode the  
 CC human secreted proteins given in AAY87064 to AAY87223. Human secreted

CC protein can have activities based on the tissues and cells the genes are  
 CC expressed in. Examples of activities include: cytostatic;  
 CC immunosuppressive; antiinflammatory; neurotropic; neuroprotective; and  
 CC antiallergic. The polynucleotides and their corresponding secreted  
 CC polypeptides are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g. by protein or gene therapy. Also pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the new  
 CC polynucleotides. Human secreted protein s and their polynucleotides can  
 CC be used for developing products for the diagnosis or treatment of cancer,  
 CC tumours, neurodegenerative disorders, developmental abnormalities and  
 CC foetal deficiencies, blood disorders, diseases of the immune system,  
 CC autoimmune diseases, hepatic and renal disease, inflammation,  
 CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,  
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,  
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,  
 CC cardiovascular disorders, reproductive disorders, gastrointestinal  
 CC disorders, respiratory disorders and metabolic disorders. The  
 CC proteins or polynucleotides can also be used as food additives or  
 CC preservatives. The proteins are also useful for identifying their  
 CC binding partners. AA298008 to AA298016 and AA87063 are sequence used in  
 CC the exemplification of the present invention.

CC Sequence 328 AA:

Query Match 31.8%; Score 1665; DB 21; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-145;  
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 703 MACRKAIVAHPLILLRLHLMIAALLGRTNHNROFROONHLSCEFLAVIGLELLOPHVF 762  
 DB 1 MACRKAIVAHPLILLRLHLMIAALLGRTNHNROFROONHLSCEFLAVIGLELLOPHVF 60  
 OY 763 RSEHOGALWDCILSFIRLLNLYRKSRRHLAFLNKFVQFIHKYITYNAPAAISFLQKHD 822  
 DB 61 RSEHOGALWDCILSFIRLLNLYRKSRRHLAFLNKFVQFIHKYITYNAPAAISFLQKHD 120  
 OY 823 PLHDLSDVDMKSLAGLSLPSRDRDTRGDLDEEESASGLPLVSVSLTPPLT 862  
 DB 121 PLHDLSDVDMKSLAGLSLPSRDRDTRGDLDEEESASGLPLVSVSLTPPLT 180  
 OY 883 AAEAPMYMRLSRGQVEDLEVLSDIDEMSRRLPILSFESNTLRLSSAECCRNLA 942  
 DB 181 AAEAPMYMRLSRGQVEDLEVLSDIDEMSRRLPILSFESNTLRLSSAECCRNLA 240  
 OY 943 FSLALRSMONSPSIAAFLPTFMYCLGSQDFEVYQALRLPEYALLQCEHAAVLLHRAF 1002  
 DB 241 FSLALRSMONSPSIAAFLPTFMYCLGSQDFEVYQALRLPEYALLQCEHAAVLLHRAF 300  
 OY 1003 LVGMVGOMPSAOISFALRIHMEAVM 1029  
 DB 301 LVGMVGOMPSAOISFALRIHMEAVM 327

RESULT 11  
 ABB59689  
 ID ABB59689 standard; Protein; 2042 AA.

XX ABB59689;  
 XX 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 5859.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.  
 OS Drosophila melanogaster.  
 PM WO200171042-A2.  
 XX 27-SEP-2001.  
 XX

PF 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, LI PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 XX N-PSDB; ABL03792.  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosure; SEQ ID NO 5859; 21pp + Sequence Listing; English.

CC Sequence 2042 AA:

Query Match 16.5%; Score 863; DB 22; Length 2042;  
 Best Local Similarity 25.4%; Pred. No. 1.5e-69;  
 Matches 280; Conservative 191; Mismatches 402; Indels 230; Gaps 32;

OY 1 MHLVVAHWYILLTLPFRADSEFQALDIWPEEKKPLPTAFLVOTSEAL-LTDDWK 59  
 DB 1084 LHINIHAFLITLTY--NSNMPESITPLIDYPPRPAPAVARLPSMPDQVOLLDPWK 1141  
 OY 60 LRMIREEVLRLVDAALDLEPQOLLFLFVSGFIPVSMKLOFLDOAVAHDPQTEQNT 119  
 DB 1142 LKMIRSSVRLIEALNDLTPDQIVLFVQNGFVMSKLLAMDVAVEGQDVKNI 1201  
 OY 120 MCKNYAHLVEVOHERGASGOTFHSLLTASLPPRRDSTPAKPKSSPEQPIQGRIRVG 179  
 DB 1202 LNKAYLAOLEIDQARGANG--HYVOALDHSHQTPPDLPKTS--VVIOEAVEI- 1254  
 OY 180 TQRLVGPEDDLAQMFLQIFPLSPDRMOSSPRALALQOLAGELARVVGSGSEVVG 239  
 DB 1255 -----DDYDS-----SDSDRPNTFLATKEVAQOTILTQPDQTESNS 1291  
 OY 240 IFVRVLOALATLTLSPH-----GGLAVNSMRSHFLACPLRLCOLCOYOR 283  
 DB 1292 DCRSLIQKLLDMLASNSNRADVYNAITTEYLVAGCVTMSRH-----ACFELR----- 1339  
 OY 284 CVPQDTGFSSFLKVLQMLQMLDSPGVESGPLRAQLRLASQASGRRLSDVRGGLRL 343  
 DB 1340 -----TFEFCMLHSDKYHILE-----NALQKMLSMF-----KHTFADSSLOK 1377  
 OY 344 AELALFRODLEVVSVRAVIATLRSGECCSVPELDS--KVLQGLEIVNSPLLELTA 401  
 DB 1378 SE--LYHESLVFRLNSREIYA-----QOFKANTALVAKRRIIRAIVO----- 1418  
 OY 402 FFSATADAASPPACKPVVVVSSILLQEEPLAGKPGADGGSLEAVRLGSSGLLVDM 461  
 DB 1419 SFDQTKDS-----KTVAKSKSDQLFH-----NGFLIDWL 1447  
 OY 462 EMIDPEVSSCPDLQL---RLFSRRKGQAOVPSFRPLTLTFTHQSSWPLHOCIRV 518  
 DB 1448 SEMDPEIVSR---QLMKERFLFSK-----SCSEFRYLLSLINHQNMWDIER-IAB 1495  
 OY 519 LIGKSRQGRDPSASLDPLMACIHVRIMQGRQRPQKRRRELVLRVGGPELISLVELI 578

Db 1496 YLFKHFHEDYATVATLNFEEALTTNPKLMGRDKYMSKNVRBDAPFMRTSELPESHFI 1555  
Qy 579 ---LAEATRSODGDTAAGSLIQARPLLLSCCGDESVAKYVEHLS(GC)QAGDSVL 634  
Db 1556 LHEGISEVKLDSKNDPLCS---RMNLEKLEKRRDLWKVEHE-----KSSVS 1605  
Qy 635 GRRCDDLQLYLRPELRVPEVLLHSEGAASSV-----CKLDGLHREFTLLA 686  
Db 1606 DYLKQVLAQVIMYPRIRKFLKP-----GKTGQAKKLQNLKCCQADKVSNNLITCLG 1658  
Qy 687 DTSRSLAENMGADASMACRIKLAVAHPLLLRHLPMAALLHGRTHLNFQEFQCNHJSC 746  
Db 1659 SLVGKKDEETSTDETLRLKLAASHPLFLROLGVLSIMGKRAQLSMKALREEHFHR 1718  
Qy 747 FLHVGLLELLQPHVF---RSEHGALMDCLSLFILLNKRKSSRHIAAFINKFQFI 802  
Db 1719 FVOILRTLELLOPTIFEEAYKNEIONTL-SCYFNFK---HNSVKEKCOMLNKFFVOML 1773  
Qy 803 HKYITYNPAALISFLQKHADPLHDLSPDNSDLVMLKSLAGLSLPSRDDRTDGLDEEG- 861  
Db 1774 QAYINYNSSALLFTEQVVGILKELAAKYTSLSGKLQVLAQVALLQHKSHSATLEDEY 1833  
Qy 862 -----EESASGSLPVSYSLFTPLIAEMA 887  
Db 1834 KYEYDLDEHFDVPSKSPVYTEDPIEVNPOTPRIDPSSSRGPLSVLTIGYSRSNYTDS 1893  
Qy 888 PY---MKRLSRGOTVEDLLEVLSDIDMSRRRPETLSFFSTNLRQMSAECCRNIAF 943  
Db 1894 PHFLDLVATIKOSNTEDEVVLGPMOELCTLSKREYFINELERLLNLTFSPSAIKSTAE 1953  
Qy 944 SLATRSKMSNP---SIAAFLPTFMYCIGSODEFVVOTALRNLPRYALLCQEHAAVLLHR 1000  
Db 1954 IILFRLKHNPNNSDINLCTINAYIQCLRDENSVAATAIDNLPMSVLLDEHAIDILTV 2013  
Qy 1001 AFLVGMGOMPSSAQSIALRL 1023  
Db 2014 AFSGLKSLNTGHQIRKVLQTL 2036

RESULT 12  
AAM95497  
ID AAM95497 standard; Protein: 161 AA.  
XX  
AC AAM95497;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen SEQ ID NO: 4155.  
XX  
KW Human; reproductive system related antigen; reproductive system disorder;  
XX  
OS Homo sapiens.  
XX  
PN WO200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01339.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226281.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234222.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.



PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249309.  
 PR 17-NOV-2000; 2000US-0249320.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0235678.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI: 2001-465570/50.  
 DR N-PSDB: AAL01467.  
 XX  
 XX Isolated nucleic acid molecule encoding a reproductive system antigen  
 PT is used in preventing, treating or ameliorating a medical condition -  
 XX  
 XX Claim 11: SEQ ID NO 4155; 1297pp + Sequence Listing: English.  
 PS  
 XX The present invention provides the protein and coding sequences of a  
 CC number of human reproductive system related antigens. These can be used  
 CC in the prevention and treatment of reproductive system disorders,  
 CC including cancer. The present sequence is a protein of the invention.  
 CC  
 XX Sequence 161 AA:  
 SQ

Query Match 10.1%; Score 530.5; DB 22; Length 161;  
 Best Local Similarity 73.1%; Pred. No. 1.5e-40;

Matches 117; Conservative 7; Mismatches 17; Indels 19; Gaps 3;  
 QY 812 AAIISFLQKHADPLHDSFSDSLVMLKSLAGLSLPRDORTRGDDEEESSAGSLP 871  
 |||  
 Db 2 AAIISFLQKHADPLHDSFSDSLVMLKSLAGLSLPRDORTRGDDEEESSAGSLP 61  
 QY 872 LVSVSLFTPLTAAEMAPYMKRLSGQTV-----DLLEVLSDIDEMSRRPETLS 921  
 |||  
 Db 62 LVSVSLFTPLTAAEMAPYMKRLSGQTVESGSPXSTPDLLESLDIDEMSRRPETLS 121  
 QY 922 FFSTNLQRLMSSADE-----CCRNLAFLRLRMQNSPSI 956  
 |||  
 Db 122 FFEDQ---PAADELRLPXKCPXPLXPALRSCRTAPAL 157  
 RESULT 13  
 ID AUAU33568 standard; Protein; 2472 AA.  
 XX  
 AC AUAU33568;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Pseudomonas aeruginosa cellular proliferation protein #12.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein;  
 XX antibiotic; antibacterial; drug design.  
 OS Pseudomonas aeruginosa.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 21-MAR-2001; 2001MO-US09180.  
 PF  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PT Yamamoto RT, Xu HH;  
 DR N-PSDB: AAS51427.  
 XX  
 XX WPI: 2001-611495/70.  
 DR  
 XX  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 XX Example 3; Seq ID No 5064; 51pp; English.  
 PS  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part



CC of the printed specification, but was obtained in electronic  
 CC format directly from WIP0 at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SO Sequence 2472 AA;

Query Match 3.0%; Score 156; DB 22; Length 2472;  
 Best Local Similarity 19.1%; Pred. No. 0.00052;  
 Matches 229; Conservative 142; Mismatches 386; Indels 440; Gaps 53;

QY 8 AMVILLITGPPRADSEFOALDIDWFPEEKPLTAFIVDTSE--EAL--LIPDM----- 57  
 DB 846 AONVSVSLPPADE-----PVDEELREVEIEEAGEVLTIGRTYIPAMKADHD 895  
 QY 58 -----LK--LRMRSEVL-----RIVVAALQDLERPOQLLFVQSF 90  
 DB 866 REALTEVRAHFTLKGSGMRAVALYIGELANSIENLFNNVDRSTIASPEV----- 947  
 QY 91 GIPVSSMSKLLQFLDOAAVANDPOTLEONIMDKNMAHLVEVOHERGASGGQTFHSILTPAS 150  
 DB 948 -----RVNDQVVALLELVEE-----FAANAQRDRDVIDLATAHALAKE 989  
 QY 151 ---LPPRDSTEARPKSSPEOPIC-----QGRIRVGTQLRVLCPEDDIAG 193  
 DB 990 PLPEPPADDGQVP--PEAGAEPSSLGNGVQAPLADAPQAALAEQSDVELLDPO----- 1043  
 QY 194 MFLQFFPLSPDRMS-----SSPRPALAQOLGELARVQSGSEVVGITVR 243  
 DB 1044 -LLEFTNEAETHLALVGLDCAARELPQVTDALQRL-----HTLKGSAHMGIL-- 1095  
 QY 244 VLQALATLSSPGALVMSHRSHFLA-----CPLLRQLQYRCVPQDTGFSLSFLTKV 298  
 DB 1096 PLAEIATFLEK-----LYKEKSNLAFDLAEALHBAEQLFRIGLEQVQAQRPLPI 1149  
 QY 299 -----LLOMLQMDSPGV-----EGPLRAQIRMLASQASAGRILSDVRCGLRIA 344  
 DB 1150 PGSDALLERIEALHOERIASLEAEVYSDAGERDPLLEAFVLEGMDILLDAEDLERWH 1209  
 QY 345 EALAFRODLEVVSSTVRAVIATLRSG-----EGCSVPRDLISKVGLLI----- 388  
 DB 1210 EHPQERQEL---SALREBLSTLDRGARHAELQVVEELQALILALDVAEEERLAVSPAF 1265  
 QY 389 --EVRSPI-----LEELLTA-----FFSATADAASPP 414  
 DB 1266 FEARQAHAELGMMDOVAAQIQTTPRPERVALQELLEAPAAEAVPTIDPSLADDP 1325  
 QY 415 -----ACKRVVVVSSLLQEEEPPLAGKPGADGSL-----EAVRLGPSSG-LIV 458  
 DB 1326 PEDEEPALPEAVYEAGAPAEETVPAAPAPAGREIDEEWVSIFLEAVDILESAGQALA 1385  
 QY 459 DWLEMLDEPVSSCPDQLRLFRSRKKGCAQVPSFRYLLTLFTHOSWFTLHCIRIV 518  
 DB 1386 QW--QAEFGAUSLSALO-RDHLTLKGARMAELAEIGD-----LAHE-----LEA 1428  
 QY 519 LIGKSRQREFDPSASLDFLMACIHVPRIWQGRDQT---POKRREELVLRV---QGP-- 569  
 DB 1429 LYEGLVDRRYQHSQPLAGLQACH-DRLAEQDQLSAGAPLADPHLIOSIRFRQGPVA 1487  
 QY 570 -----ELISVELLAEETRSQDGTACSLQIARPLLLSCCGDESVRYKTEHUS 623  
 DB 1488 EAATPGEAESPVLEIYAPA---VEEPAPAAAEFERDELEIFL---EEGFDIIDLSAA 1541  
 QY 624 GCIQMGDSV----- 633  
 DB 1542 AALQRMMDVNTTELEALQRLHLTKGARMATIGEIDLAHELFTEGLCGRLRAS 1601  
 QY 634 -----LGRRCRDLQLQY-----LGR-----PELRVVPPEVL---LH 662  
 DB 1602 PALFGILQCHDELAEMLAEVAGHRTLPDQGLALAEIRLRASDPDQLSVPSVSLKPLA 1661  
 QY 663 SGAASSSVCKDGLHRTITLAD-----TSRRLDENNGADASNAACKRLAVA 711  
 DB 1662 AAGAAADESEIIDFLEADLDLENLELALGRWDGNGDAQPLDD----- 1706

QY 712 HPLLRLHPLMIAALLHGRTHLNFQEFROQNLSCFLHVLGLLELOPHVERSEHOGALW 771  
 DB 1707 --LRLRLHLTKGARLACQTEIG-----NLAHLEQHLTAAQOOGAPW 1747  
 QY 772 -DCLLSFIRLLNLYRKSRSNHLAFLNKEVQFIHKYITYNAPPAISFLQKHADPLHDSFD 830  
 DB 1748 PDSL-----LDAQSGLBGLQROVDLRLRLAE 1775  
 QY 831 NSDLYMKSLLAG-----LSLPSRDDTRD---GLDEEGEBSAGSLPVYSILTFP 880  
 DB 1776 DDE-----AGERPEAQLVQADD-TDRAVASALAEITRLAPAGA--TMAEAAP 1824  
 QY 881 LTAEMAPYMKRLSRQGVEDLEVLSDIDEMSRRP-EILSFFSTNLORLMSAAE 936  
 DB 1825 AAPATTLFVVKRAQEAQ-----EASRRAPQELVAVPALLENLVNLAGE 1870

## RESULT 14

AAM52358  
 ID AAM52358 standard; Protein; 1217 AA.

AC AAM52358;

DT 25-JAN-2002 (first entry)

DE Putative TBP interactive protein.

KW Gemlinivirus; plant; viral infection; transgenic plant;

KW tomato yellow leaf curl virus.

OS Arabidopsis thaliana.

PN FR2806095-A1.

PD 14-SEP-2001.

PF 10-MAR-2000; 2000FR-0003140.

PR 10-MAR-2000; 2000FR-0003140.

PA (GENT-) GENTECH SARL.

PI Bejarano ER, Castillo GA, Colinet D, Donoso CI, Iniesta JR;

PI Grevesse C, Hericourt F;

XX WPI; 2001-628275/73.

DR N-PSDB; ABA01228.

PT New polynucleotides for producing transgenic plants resistant to

PT geminivirus infection comprising polynucleotides encoding proteins

PT which interact with at least one of the products of the geminivirus

PT genome

PS Claim 4; Pages 44-48; 106pp; French.

XX The present invention relates to coding sequences encoding proteins which

XX interact with at least one of the six products of the geminivirus genome

CC necessary for infection of a plant by the virus. The present sequence is

CC one such protein. The coding sequences are useful for producing

CC transgenic plants resistant to geminivirus infection, particularly tomato

CC yellow leaf curl virus.  
 SO Sequence 1217 AA;

Query Match 2.8%; Score 149; DB 22; Length 1217;

Best Local Similarity 20.4%; Pred. No. 0.00074;

Matches 196; Conservative 137; Mismatches 334; Indels 292; Gaps 49;

QY 46 DTSEALILPDMILKLMISSEVLRVDAALQDLERPOQLLFPQSGFPIPVSSMSKLLQFLD 105  
 DB 409 DTDESS---PKML---LKQEVSKIYKSIINQLREKSVTKVGAFFSV---LRELVVLP 457

QY 106 QAVAHDPOTL-----EONINDKNYMAHL-VEVQHERGASGQOTHSILITASLPPRRODSTEA 160  
 DB 458 DCIADHIGSLVPGIERALANKSTSNKITE-----ALVFETKIVLASHAP----- 501  
 QY 161 P-----KPKSSPE-PIGQGRIRV-----GTOLRVIGP----- 167  
 DB 502 PVFHPYIKALSSVYLAAGVGRYKVTAEALRVGELRVVRFSTAGNGFDFKPPVHETYN 561  
 QY 188 -----EDDLAGMFLQIFPLSPPRMOSSSPRPVALALQALGOELARVY 231  
 DB 562 AIMSRLTNDQDQEVKCAITCMGLVISTGDDLRALPSCLEPV---LVDRMGNEIRLT 618  
 QY 232 QGSEPEVPGITVRVQLATILSSPHGALVMSHRSHFACPLRLQCLQYRCVPODPTG 291  
 DB 619 -----AVKAFSVYATSP-----LH-----INLSCVLDHLIAELTGF 649  
 QY 292 SSLEKVLQMLQMLDSPGVEGGPRLRAQLMLASQASA-GRR-L-SDVRGGLRLAEALAF 349  
 DB 650 LKRAKRVLR-----ATLITMTLVTAIGDKIGSEAYEVILVELSLIS 693  
 QY 350 RODLEVSVSTRAVIAVLNRSGEQSVEPDLI--SKVL-QGLIEVRSFHEEL---LTAF 402  
 DB 694 VSDLHMTALALE-LCCTLMGKSCSENI SLAVRNKVLPAQLTLVKSPLLGQALLDQKF 752  
 QY 403 FSATA-DAASPFPACKRVVYVSSIL-LQEEPILAGKP-----GADGC 443  
 DB 753 FEALVYHANTSF-----YTLLESILSCAKPSPOSGVPKQALYSIACVAVLCLAGDN 807  
 QY 444 SLFAVRLGSPSSGLLVLMLEMDPEVWSSCPDLQRLRL---FSRRKG-KCOAQVPSRPY 498  
 DB 808 CSSTVAK-----LMELTKDSDGNSAKQHLALLSLGEIGRRKLSHAGET----- 854  
 QY 499 LTLTTHQSSWPLTHOCITRVLCKSRQRPDSALDFLMACIHVPRIMOGRDOFTPOKR 558  
 DB 855 -IYIESFOSPFEEKISNAYSALGNINAVNL--SNYLPEFI-----LDQIDMOQK 899  
 QY 559 REELVRYVQPELISLVELLAEFTRFSGODGNAACSLIQARLPLILSCCGGDESRYK 618  
 DB 900 KQYTLHL-----SLKEVIVRO---SVDKADFONSSVERKTLALFNHCESEEGVRN 948  
 QY 619 TEHLSCGCIQMGDSVYLRRCRDLLQYLQRPETLRVVPVLLHSEGAASSVC----- 672  
 DB 949 VARCLG-----KWLLEPEKLVPAALQVRTSPAFRAVTVAVKY 989  
 QY 673 -----KLDGL-----IHRITTLADTSDSRALENRGADSMACRKIAVAH--PLILRH 719  
 DB 990 SYVEREKDEIETFPQISSFLMLIKD-GDRHV--RRAAVALS---TFAHYKFNILIKGL 1042  
 QY 720 LPMIALLLHGRTHLNQEPFQONHLSCEFLVH--GILELLQPHYFRSEHOGALMDCLLSF 777  
 DB 1043 LPELLEPLLDQYVYK-KELIRIVDLGPFKHAVDDGL-----ELKRAAECEVFTL 1090  
 QY 778 IRLLLNRKSSRHAAAFINKVQ-----FIHKYITYNAPAI-SFLQKHADPLH- 825  
 DB 1091 VDSCLQDVNNSFIVPLKSGLEDHYDKMLCHILSLADKCPSAVLAVIDSLVEFLHK 1150  
 QY 826 DLSFD-----NSDLVMLKSLLAGLSLPSRDRTDGLDEGESESSAGSPL 872  
 DB 1151 TISFKPKODAVKQEHDRNED--MIRSAIRAISSLDIRINGVDYSHKFGKMGDMKRSVPL 1207

RESULT 15  
 AA85660  
 ID AA85660 standard; Protein; 1519 AA.  
 AC AA85660;  
 XX  
 DT 12-FEB-2001 (first entry)  
 XX  
 XX Human tyrosine kinase substrate tks107/GRUB protein sequence.  
 XX Tyrosine kinase substrate: Tks: Tks 107; GRUB: Tks 113; Tks 118; Dresch;  
 KM Tks 202; Src: rheumatoid arthritis; atherosclerosis; stroke; cancer.

KW autoimmune disorder; organ transplantation; myocardial infarction;  
 KW cardiomyopathy; renal failure; neurodegenerative disorder;  
 KW cardiovascular disorder; inflammatory bowel disease; multiple sclerosis;  
 KW asthma; osteoarthritis; psoriasis; rhinitis.  
 OS Homo sapiens.  
 PN WO200061750-A2.  
 PD 19-OCT-2000.  
 XX 06-APR-2000; 2000MO-US09277.  
 PF 09-APR-1999; 9905-0128492.  
 PR (SUGEN-) SUGEN INC.  
 PA Phan H, Courtneidge SA;  
 XX WPI, 2000-679486/66.  
 DR N-PSDB; AAC66076.  
 XX  
 PT Novel tyrosine kinase substrate polypeptides and polynucleotides, used  
 PT to diagnose and treat diseases such as rheumatoid arthritis,  
 PT atherosclerosis and cancer -  
 PS Claim 2; Fig 5; 120pp; English.  
 XX  
 CC This invention relates to isolated, enriched or purified nucleic acid  
 CC molecules represented by sequences AAC66076-066079 which encode tyrosine  
 CC kinase substrate (Tks) Tks 107/GRUB, Tks 113, Tks 118/Dresch or Tks 202  
 CC represented by sequences AA85660-165663. The Tks proteins are  
 CC substrates for cytoplasmic tyrosine kinase Src. The invention encompasses  
 CC probe sequences, recombinant cells, antibodies specific for the Tks  
 CC proteins, and hybridomas producing the antibodies. The Tks proteins  
 CC exhibit antirheumatic, antiarthritic, antiarteriosclerotic;  
 CC immunosuppressive; cardiact; cerebroprotective; neuroprotective;  
 CC cytoskeletal; antiinflammatory; osteopathic; antipsoriatic; antiallergic;  
 CC and antiautismic activities. The tyrosine kinase substrates are used to  
 CC treat and diagnose diseases or disorders selected from rheumatoid  
 CC arthritis, atherosclerosis, autoimmune disorders, organ  
 CC transplantation, myocardial infarction, cardiomyopathies, stroke, renal  
 CC failure, oxidative stress-related neurodegenerative disorders,  
 CC cardiovascular disorders, cancer, and immune-related disorders selected  
 CC from chronic inflammatory bowel disease, chronic inflammatory pelvic  
 CC disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, and  
 CC rhinitis. The proteins can also be used to identify modulators of its  
 CC activity. DNA sequences encoding the proteins may be used in gene therapy  
 CC methods, and to produce transgenic animals.  
 CC  
 SQ Sequence 1519 AA:  
 Query Match 2.7%; Score 143; DB 21; Length 1519;  
 Best Local Similarity 19.7%; Pred. No. 0.0038;  
 Matches 190; Conservative 87; Mismatches 284; Indels 402; Gaps 41;  
 QY 12 LTLGPRRADSEFOALLDIFPEKEPLPTAFVDTG---EALLLPDMTKLRIRSEVL 68  
 DB 549 LRLTTPP-----CPPEPSPSRDLNTLTHYLSLRD----- 582  
 QY 69 RVDAAIODELEPOQLLFVOSFGIPVSMKSLQFLQQA-----VAHDPOTLE-- 116  
 DB 583 -----LQTLGLSVLIDLKRAAPLIPALIPALISOLQDSGPPVQLLIHDDLPLELC 636  
 QY 117 -----ONIMDKNYM-----AHLVEYQHE-----RGASGC--Q 141  
 DB 637 GFGAEVLSNDLKRVAKPELOMELGHRDPSSHVEVRLHQEVVRLCRICQGLVSGVRQ 696  
 QY 142 TFSLSLTALEPPRRDSTEAPKSPSPQPIGQGRIRVGTQLRVYLGPEDDLQAGMFLQIFPL 201  
 DB 697 AIELLEGAAPEPEEEAVGEMKPIQ-----KVLADEPRILTAORD--GAILMLRLRS 744  
 QY 202 SPDRMOSSSPRPVALALQALGOELARVVGSPFEVPGITVRVQLATILSSPHGALV 261

```
Db 745 TPSSKLECOGPATLYQVEYDEAIIHQ-LVR-----LSNLH----- 776
QY 262 MSMHSHFLACPLRLQLOQYQRCVPQDGFSSLFKVLQMLQWLDSRGVE-----G 313
Db 777 -----VOQGEQROC-----LRLQOVLQWLSPGEEQLASFAMPG 811
QY 314 GPLRA-----QLRMLASQASAGRRLSDVR----- 337
Db 812 DTLSLQETELRFRAFSAEVQERLQAAREALALEENATSQKVLDFEQRLQVEESGLHRA 871
QY 338 -----GGLRLA-----EALAFRODLEVVSST---YRAVI----- 364
Db 872 LRLQFFQOAHHEWVEGFARLAGAGRGREAVLALALARRAPEPSAGTFOEMRALALDLSG 931
QY 365 -ATLRSGEQCSE-PDLISKVYLQGLIEVRSPILELLTAFFSATADAA-----SP 412
Db 932 PAALREMGRCQARCOELERRIOHILGEASPR-----GYRRRRADGASSGAGQWGPSP 985
QY 413 FPAKPVVVSLLQEEEPILAGKPGADGSL-----EAVRLGPSS-----GIL 457
Db 986 SPS-----LSSLIPSS---PGPRPASHCSLACGEDYEEGPELAPAEGRPPRAVL 1036
QY 458 VDWLEMLDPEVVSQPDQLRLLFSSRRKGQAOVPSFRPYLLTLFTHQSSWPTLHOCIR 517
Db 1037 IRGLEVTSTEVVD-----RTCSPREH----- 1057
QY 518 VILGSRQORPDSALDFLWACHIVPRIMQGRDORTPOKRRELVLYVQ----- 567
Db 1058 VILGRAR-----GPDGPMG-VGAPRMERKRISIAOQRLVSELJACEQDYVATLSEPV 1108
QY 568 ---GPGLISLVELLIAEAFRSQDGTAACSLIQARL-----PLLS 606
Db 1109 PPPGPPL-----TPELRGTMAAALSARELRSFHRTFLRELQGCATHPL--- 1153
QY 607 CCCGDDESVRKVTREHLSCGIQWGDVILGRRCRDILLQYLQRPPLRVVPPEVLLHSEGA 666
Db 1154 -----RIGACFLRHGDOF-----SLYAQYVKHRHKTENGLAISPLSKGS 1193
QY 667 ASSSVCKLDGL-----IHRFTTLADTSDSRALENRGADASMACKLAVANHPILL---L 717
Db 1194 MEAGPYLPRALOQPLEQLFRIGRLLE-----LIREAGPELSSCCRALGAANVQLREDEA 1248
QY 718 RHLPMIAALLHGRTHLNFQEFROQNHLSCLFVLGLLELLOPHVFRSEHOGALMDCLLSF 777
Db 1249 RGRDILAVEAVRGCEDLKEQGQLLHRDPFTYICGRKKCIR-HVFLFEH-----LLLF 1300
QY 778 IRL 780
Db 1301 SKL 1303
```

Search completed: May 7, 2003, 19:40:57  
Job time : 55 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:39:33 ; Search time 27 Seconds  
(without alignments)  
3663.790 Million cell updates/sec

Title: US-09-929-769-7

Perfect score: 5241

Sequence: 1 MHILVHAMVILTLTGPPRA.....MDPSAQSIALRIILHMEAVM 1029

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3244	61.9	640	2 T08758	hypothetical prote
2	1240.5	23.7	310	2 T46429	hypothetical prote
3	156	3.0	2472	2 E83594	still frameshift p
4	154	2.9	1638	2 T30313	chemotaxis protein
5	149	2.8	1217	2 T00607	hypothetical prote
6	147.5	2.6	878	1 RRSIB	RNA-directed RNA p
7	134.5	2.5	2535	2 T04824	hypothetical prote
8	131.5	2.5	882	2 AE0119	Clp Arpase (import
9	131.5	2.5	1386	2 T00257	hypothetical prote
10	129	2.5	1626	2 A75613	hypothetical prote
11	128	2.4	1012	2 I53172	RAE-28 - mouse
12	128	2.4	1021	2 H75423	hypothetical prote
13	126	2.4	880	2 AE0179	probable Arpase ch
14	125.5	2.4	1196	2 T13057	KIAA0729 protein -
15	125.5	2.4	1211	2 C83110	probable exonuclea
16	125	2.4	949	2 T44577	hypothetical prote
17	123	2.3	939	2 G75481	sensory box/GGDEF
18	122.5	2.3	738	2 A53542	breifeldin A-sensit
19	122.5	2.3	1366	2 T35985	probable large pro
20	121	2.3	1134	1 A35955	meta-vinculin - ch
21	121	2.3	1135	1 A39997	meta-vinculin - ch
22	121	2.3	1239	2 S74355	hypothetical prote
23	120.5	2.3	882	2 T43250	splindle pole body-
24	120.5	2.3	1111	2 T01239	hypothetical prote
25	120	2.3	555	2 H83043	hypothetical prote
26	120	2.3	2698	2 B96671	similar to transla
27	120	2.3	3149	1 OQBE8	Bip1 protein - hu
28	119.5	2.3	818	2 S62790	mismatch DNA recog
29	119.5	2.3	2048	2 C84609	hypothetical prote

30	119.5	2.3	2895	2 H85362	hypothetical prote
31	118.5	2.3	920	2 H82302	ATP-dependent heli
32	118.5	2.3	814	2 T17233	hypothetical prote
33	118.5	2.3	1237	2 A31789	band 3-related pro
34	118.5	2.3	1237	2 A56764	band 3-related pro
35	118	2.3	1136	2 H9564	F508.33 [imported]
36	117.5	2.2	2388	2 JE0271	beta spectrin, bet
37	117	2.2	1066	2 T10108	vinculin - mouse
38	117	2.2	2442	2 T08621	centrosome associa
39	116.5	2.2	1234	2 A34911	band 3-related pro
40	116.5	2.2	1538	2 B90924	probable ATP-depen
41	116.5	2.2	1538	2 F85772	ATP-dependent heli
42	116.5	2.2	2541	2 S11661	tailin - mouse
43	116.5	2.2	2777	2 D96746	hypothetical prote
44	116	2.2	465	1 WMBE28	UL38 protein - hum
45	116	2.2	5032	1 A35041	ryanodine receptor

#### ALIGNMENTS

##### RESULT 1

T08758 hypothetical protein DKFZp586J0619.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999

C:Accession: T08758

R:Wandut, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z16471

A:Accession: T08758

A:Molecule type: mRNA

A:Residues: 1-640 <MAN>

A:Cross-references: EMBL:AL050110

A:Experimental source: adult uterus; clone DKFZp586J0619.

A:Genetics:

A>Note: DKFZp586J0619.1

Query Match	Score 3244;	DB 2;	Length 640;
Best Local Similarity	100.0%;	Pred. No. 4e-222;	
Matches 633;	Conservative	0;	Mismatches 0;
		Indels 0;	Gaps 0;
QY	397	ELTTAFSATAADAAFPACKPVVVSSLLDDEEPLAGCKPGADGSLAAVHLGSSGL	456
DB	8	ELTTAFSATAADAAFPACKPVVVSSLLDDEEPLAGCKPGADGSLAAVHLGSSGL	67
QY	457	LYDMLMDPEVYSSCPDLQLRLPSRRKKGQAOVPSRPYLLTFTQSSPPTLHOCI	516
DB	68	LYDMLMDPEVYSSCPDLQLRLPSRRKKGQAOVPSRPYLLTFTQSSPPTLHOCI	127
QY	517	RYVLGKSRQRFDPASLDFLMACIHVPRIWGRDQRTPOKREELVLRVSGHELISYE	576
DB	128	RYVLGKSRQRFDPASLDFLMACIHVPRIWGRDQRTPOKREELVLRVSGHELISYE	187
QY	577	LILAEETRSDQDPAACSLIQARPLLLSCCGDDESVKYTEHLISGCIQMGDSVLGR	636
DB	188	LILAEETRSDQDPAACSLIQARPLLLSCCGDDESVKYTEHLISGCIQMGDSVLGR	247
QY	637	RCRDLLQIYDQRPRLRVPEVYLHSEGAASSVYKLDGLIHRFTLLADVSDSALEN	696
DB	248	RCRDLLQIYDQRPRLRVPEVYLHSEGAASSVYKLDGLIHRFTLLADVSDSALEN	307
QY	697	RCADASMACRKLAVAPHLILRLPMTALILGRTPLNFQEFROQNHLSCTLHVGLLEL	756
DB	308	RCADASMACRKLAVAPHLILRLPMTALILGRTPLNFQEFROQNHLSCTLHVGLLEL	367
QY	757	LQPHVFRSEHQALWDCLLSFIRLLINLYRKSSRHLLAFINKEVQFTIKYITYNAPAISE	816
DB	368	LQPHVFRSEHQALWDCLLSFIRLLINLYRKSSRHLLAFINKEVQFTIKYITYNAPAISE	427
QY	817	LQKHADPLHDLSPFNSDLYMLKSLAGLSLPRDRDQDGLDEEGEESASGLPVYSV	876
DB	428	LQKHADPLHDLSPFNSDLYMLKSLAGLSLPRDRDQDGLDEEGEESASGLPVYSV	487

QY 877 LFTPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRREPELISFSTNLQRLMSSAE 936  
 DB 488 LFTPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRREPELISFSTNLQRLMSSAE 547  
 QY 937 CCRRLAFLSLRLSKMNSPSTAAAFLEPTFMVCLGSQDFEVYQVOTALRNLPEVALLCOEHAAY 996  
 DB 548 CCRRLAFLSLRLSKMNSPSTAAAFLEPTFMVCLGSQDFEVYQVOTALRNLPEVALLCOEHAAY 607  
 QY 997 LHRRAFLVGMVGMQMPDPSAQISEALRIIMEAVM 1029  
 DB 608 LHRRAFLVGMVGMQMPDPSAQISEALRIIMEAVM 640

RESULT 2  
 T46429  
 hypothetical protein DKFp434C0126.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46429  
 R:Ansoyge, W.; Wilkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: 223028  
 A:Accession: T46429  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-310 <AAA>  
 A:Cross-references: EMBL:AL137358  
 A:Experimental source: adult testis; clone DKFp434C0126  
 C:Genetics:  
 A:Note: DKFp434C0126.1

Query Match 23.7%; Score 1240.5; DB 2; Length 310;  
 Best Local Similarity 92.6%; Pred. No. 2; Le-80;  
 Matches 249; Conservative 2; Mismatches 7; Indels 11; Gaps 1;

QY 756 LDPHVRSEHOGALMPCILSFTRLILNKRSSPHLAFTNKPFQFIHKYTYNAPAAIS 815  
 DB 1 LDPHVRSEHOGALMPCILSFTRLILNKRSSPHLAFTNKPFQFIHKYTYNAPAAIS 60  
 QY 816 FLQKHADPLDLSPNSDLVMLKSLAGLSLPSRDRTDRGLDEGEESAGSLPLVSV 875  
 DB 61 FLQKHADPLDLSPNSDLVMLKSLAGLSLPSRDRTDRGLDEGEESAGSLPLVSV 120  
 QY 876 SLEPTPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRREPELISFSTNLQRLMSSAE 935  
 DB 121 SLEPTPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRREPELISFSTNLQRLMSSAE 180  
 QY 936 ECCRNLAFLSLRLSKMNSPSTAAAFLEPTFMVCLGSQDFEVYQVOTALRNLPEVALLCOEHAAY 995  
 DB 181 ECCRNLAFLSLRLSKMNSPSTAAAFLEPTFMVCLGSQDFEVYQVOTALRNLPEVALLCOEHAAY 240  
 QY 996 VLLHRAFLVGMVGMQMPDPSAQISEALRIIMEAVM 1029  
 DB 241 VLLHRAFLVGMVGMQMPDPSAQISEALRIIMEAVM 640

RESULT 3  
 E83594  
 still frameshift probable component of chemotactic signal transduction system PA0413 [1]  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: E83594  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mlczoguchl, S.D.; Warren, P.; Hickey, M.J.; B.  
 adman, S.; Yvan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lm,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MOID:20437337; PMID:10984043  
 A:Accession: E83594  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2472 <STO>

A:Cross-references: GB:AE004478; GB:AE004091; NID:99946261; PIDN:AA03802.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA0413

Query Match 3.0%; Score 156; DB 2; Length 2472;  
 Best Local Similarity 19.1%; Pred. No. 0.042;  
 Matches 229; Conservative 142; Mismatches 386; Indels 440; Gaps 53;

QY 8 AMVLLTLPGRADDSFQALDLDPPEKPLPAFLVDNSE--EAL--LLPDV----- 57  
 DB 846 AQNVPSVLLPPADE-----FYDELRVFLIEEAGEVLEIGRIIPANKADHD 895  
 QY 58 -----LK--LRMRSEVL-----RLVDAALDLEPQQLLFVQSF 90  
 DB 896 REALTEVRAFFHTLKGSGRWRAVLIGELAWSIENLFNRYLDRISAEPVQ----- 947  
 QY 91 GIPVSSKLLQFLDQAVANDPQTLQONIMDKNMALVQVHGRGSGGQTFSLTAS 150  
 DB 948 -----RVVDQVALLPELVEE-----FAANAQRQRDVDLLAATAHALAKE 989  
 QY 151 ----LPPRDSTEARPKSPSEPIG-----OGRLIVGQLRYLAGEPDLAQ 193  
 DB 990 PLRPPAPADGGVP--PAGAGQPSLNGVQAPPLADPAQAANAQSDVELDPQ----- 1043  
 QY 194 MFLQIFPLSPDPRMOS-----SSPRVALALQALQELARVQGSPEVPGITVR 243  
 DB 1044 -LLEIFTEAETHLEALVGLADCARELQPVYDIALQAL-----HTLKGSAHAGIL-- 1095  
 QY 244 VLQALATLLSPHGALVMSHRSFLA---CRLRLQCOVQRCVPQDTGFSFLKY 298  
 DB 1096 PLAIETAPLEK-----LVAEYKSNLAFDLRELLHDAEQLEFRIQEDVGAORPLNPI 1149  
 QY 299 ----LLOMLQWLDSPGV-----EGPLRAQLMMLASQASGRSLDVRGCLR 344  
 DB 1150 PGSDALLERIEALHGERIASLEARYSDAGERRPLIEAFVYGMQILDAEDLLRWH 1209  
 QY 345 EALAFRODLEVYSTVAIVATLRSQ-----EQCSVEPLIKVYLOGL----- 388  
 DB 1210 EHPEROEL---SALREELTLDRGARHMLPOVEELCALALAYAVEGRLAVSPAF 1265  
 QY 389 --EVRSPH-----LEELTA-----FFSATDAASPP 414  
 DB 1266 FEEARQAHLEALIGMDQVAAAGLOVTPRPERYALQELLEAPAAVFIPEDESIGADDF 1325  
 QY 415 ----ACKPVVVVSSLLQEEPLAGCKPGADGSL-----BAVRLGPSSG--LLV 458  
 DB 1326 PEDEPALPEAVYEAGAPAEETVPAPAPAPGRELDEBWSIFLEAVDILBSAGALA 1385  
 QY 459 DWLEMLDPEVYSSCPDQLRLSRRKKGQAOVPSRPYLLTLFTQOSSWPTLHCIRV 518  
 DB 1386 QW--QAEPGALSSLSAQ--RDLHTLKGARMAETAEIGD-----LAHE-----LEA 1428  
 QY 519 LLKSGRORPDSASLDFLWACIHVPRTWGRDRT---DOKREELVLRV---OGP-- 569  
 DB 1429 LYEGLVDRRIQHSQPLGLQIACH--DRILAEOLDLSAGQPLADPHDILQSTRRRQGPVA 1487  
 QY 570 -----ELISVELILAEATRSQDGTAAACSLIOARLPLLSCCGDDESVRKVTHEL 623  
 DB 1488 EAATPGAESEVVELVAP---VEEPAPAAEAEEEDPEVEFL--EEGFIILDSAA 1541  
 QY 624 GCIOQMGDSV----- 633  
 DB 1542 AALORWMDVDVNTIELEALORDHTLTKGARMAEIGELGLAHLEFLYEGLCGRRLAS 1601  
 QY 634 ----LORRCRDLLLOLY-----LQR-----BELAVPYEVL-----LH 662  
 DB 1602 PALFGLQORCHDELAELVAVRGHRTLPGQALIAETIRSRSDPEQLSVPTSVSLKPLA 1661  
 QY 663 SEGAASSVCKLDLIRFTILLAD-----TSDRALENRGADASMACKRLAVA 711  
 DB 1662 AKGAADSESLDIFLEADDLLENLELALGRMDGNGADQPLDD----- 1706

QY 712 HPLLLRLPMIALALHGRTHLNFQFROONHLSCLFHLVGLLELLOPHVFRSEHGALM 771  
 Db 1707 --LRLHLTKGKARLAGTELG-----NLAHDEQLHTDAQOOGARW 1747  
 QY 772 -DCLLSFIRLLINRKSSRHIAAFINKFVQFIHKYITYNAPPAISFLQKHADPLHDSFD 830  
 Db 1748 PDSLL-----LDAQGLSEGLQROVQLRERLAE 1775  
 QY 831 NSDLVMTKSLLAG-----LSIPSRDDRTDR-----GLDEGEESAGSLPIVSVSLFTR 880  
 Db 1776 DDE-----AGERPEPAQALVQADD--TDRAVASALAEITRLAPAGA--TMAEAAP 1824  
 QY 881 LTAEMAPYMKRLSRGQVEDLEVLSDIDEMRRRP--EILSFFSTNLORLMSAAE 936  
 Db 1825 AAPATTLPEVVRKAQEAQ-----EASRRAPQELVVPAPALLEINVLNAGE 1870

## RESULT 4

130313  
 chemotaxis protein homolog - Pseudomonas aeruginosa  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T30313  
 R:Whitchurch, C.B.; Young, M.D.; Hobbs, M.; Matlick, J.S.  
 submitted to the EMBL Data Library, November 1996  
 A:Description: Pseudomonas aeruginosa chemotactic transduction genes pilL, chpA chpB and  
 A:Reference number: 220819  
 A:Accession: T30313  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1638 <WHI>  
 A:Cross-references: EMBL:U79580; NID:g3241967; PID:g3241969; PIDN:ACC23931.1  
 C:Genetics:  
 A:Note: chpA

Query Match 2.9%; Score 154; DB 2; Length 1638;  
 Best Local Similarity 19.1%; Pred. No. 0.032;  
 Matches 229; Conservative 141; Mismatches 387; Indels 440; Gaps 53;

QY 8 AMVILLITGPPRADDSERQALLDIFPEEKPIPTAIVDTSE--LAL--LLEDM----- 57  
 Db 12 AONVPSLLPPADE-----PYDELRVFIIEAGVLEFTRGLYRLPAMKADHD 61  
 QY 58 -----LK--LMIRSEVL-----RLVPAALODLEPOOLLFVQSF 90  
 Db 62 REALTEVRAFFHTLKGSGMVRALVIGELAWSIENLFNVLDIRSTIASPPVQ----- 113  
 QY 91 GIPVSSMSKLLQFLDQAVAHDPQLEQNTMDKNMAHLVEVQHERGASGGQTFHSLTAS 150  
 Db 114 -----RVVDQVVALPELVEE-----FAANAQRQRDVDLLAATAHALAKE 155  
 QY 151 --LPPRHSTAPRKSPPEPIC-----QGRVGTQLRVLGPEDDLAG 193  
 Db 156 PLPEPPADPDGGVP--PEAGAEQPPSSLDNGVQAPPLADAPQAAAEASDVLELDPQ----- 209  
 QY 194 MFLQIFPLSPDPRMOS-----SSRPALALQALQGLARVVGSPFEPQITVR 243  
 Db 210 -LLEIFTNAETHLEALVGFADCARELPOVYTDALQRL-----HTLKSAMHAGIL-- 261  
 QY 244 VLQALATLLISPHGALVMSHRSHFLA-----CPLLROLQYQRCVPDQTFSSSLFKV 298  
 Db 262 PLAEIATPLEK-----LYKEYKSNLLAEFLDREAELIHDAEQLFRIGLQVGAQRLNPI 315  
 QY 299 -----LLQMLQWLDSRGV-----EGGPLRAQLRMLASQASAGRRSLDVKGGILLRLA 344  
 Db 316 PGSDALLETLELHQRASLAEARYSDGERRDPLLEAFVLEGMDILLDEDLLEIRWH 375  
 QY 345 EALAFRODEEVSVTRAVIATLRSG-----EGCSVEPDLISKVLOGLI----- 388  
 Db 376 EHPQERQEL-----SALRELSLTDRCARHAELEPQVELCOALLALYDAVEBGLAASPAF 431  
 QY 389 --EVRSPP-----LEELLTA-----FFSATADAASPP 414

Db 432 FFEARQAHALLIGMMDQVAGLQVTPRPERVAALQELLEAPAAAEVFPIDPESLGADPP 491  
 QY 415 -----ACKRVVYVSSLLDEEERPLAGKRGAGGSL-----EARRLGPSSG--LLV 458  
 Db 492 PEDEBPALPEAAVEEAGAPAEETVPAAPAPAGREIDEEMVASIFLEAVDFLESAGQALA 551  
 QY 459 DWLEMDPEVSSCPDLQRLLEFSRRKKGQAOVSFRFYLLTLTHQSSWPTLQICIRV 518  
 Db 552 QW--QAEFALSSLSLQ--RLHTLKGARMAEIAEIGD-----LAHE-----LEA 594  
 QY 519 LIGKSHRQRPDSASIDFLMACIHVPRIWQGDQRT--PQRREELVLRV---QGP-- 569  
 Db 595 IYEGLVDRRYQSPQLAGLQACH--DRLAEQLDQLSAGQPLADPHLIQISIRFRKQPYA 653  
 QY 570 -----ELISVLELLIAEFTSSQGDPTACSLIQARLPLLLSCCGDDESRYKTYEHS 623  
 Db 654 EATPGEASEPVEELVAP--VEEPADAPAEFERDELVEIFL--EEGHLIDSAA 707  
 QY 624 GCIQWGDV----- 633  
 Db 708 AALQRMMDVDVNTIELAQRLHTLKGARMAEIGELIDLAHELEFLYEGLCGGRLAS 767  
 QY 634 -----IGRRCDLLQLY-----LQR--PELRVPEVL--LH 662  
 Db 768 PALFGLHQCHDELAEMLFAVGHRTLPDQALAEIRLRSPDPDQLSVPTSVSLKPLA 827  
 QY 663 SGGAASSVCXKIDGLHRTITLAD-----TDSRLAENNGAASMAKRLAFA 711  
 Db 828 AKGAADSEELIDFLEADLLENLELALGRWDGNGDAQPLDD----- 872  
 QY 712 HPLLLRLPMIALALHGRTHLNFQFROONHLSCLFHLVGLLELLOPHVFRSEHGALM 771  
 Db 873 --LRLHLTKGKARLAGTELG-----NLAHDEQLHTDAQOOGARW 913  
 QY 772 -DCLLSFIRLLINRKSSRHIAAFINKFVQFIHKYITYNAPPAISFLQKHADPLHDSFD 830  
 Db 914 PDSLL-----LDAQGLSEGLQROVQLRERLAE 941  
 QY 831 NSDLVMTKSLLAG-----LSIPSRDDRTDR-----GLDEGEESAGSLPIVSVSLFTR 880  
 Db 942 DDE-----AGERPEPAQALVQADD--TDRAVASALAEITRLAPAGA--TMAEAAP 990  
 QY 881 LTAEMAPYMKRLSRGQVEDLEVLSDIDEMRRRP--EILSFFSTNLORLMSAAE 936  
 Db 991 AAPATTLPEVVRKAQEAQ-----EASRRAPQELVVPAPALLEINVLNAGE 1036

## RESULT 5

T00607  
 hypothetical protein At2g02560 [imported] - Arabidopsis thaliana  
 M:Alternate names: hypothetical protein T8K22.14  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 02-Feb-2001  
 C:Accession: T00607; B84438  
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K  
 submitted to the EMBL Data Library, June 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC T8K22 genomic sequence.  
 A:Reference number: 214192  
 A:Accession: T00607  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1217 <ROU>  
 A:Cross-references: EMBL:NC004136; NID:g3184270; PID:g3184283  
 A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujita, C.Y  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
 euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: B84438  
 A:Accession: B84438  
 A:Status: preliminary  
 A:Molecule type: DNA





Db 588 --DKRFLCSAAYPKGVENKSL-----KSKVGIQQAAYVRYEALRLVGMNYPPLNK 638

Qy 607 CCCGDESVRYKTE-----HLSCIOQMGD-SVIGRRCRDLLLOLYLOR---DELRLVPV 657

Db 639 ACKNNAARHLLEAKGPFLLDEFLEAMSELSEFGTEFEGFNIKLIVTRENLAFLNKVP 697

RESULT 7

T04824

hypothetical protein F10M23.350 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999

C:Accession: T04824

R:Bevan, M.; Lecharny, A.; Chedior, F.; Kravitzky, M.; Kreis, M.; Hohnsels, J.; Mewes, H

submitted to the Protein Sequence Database, February 1999

A:Reference number: 215385

A:Accession: T04824

A:Molecule type: DNA

A:Residues: 1-2535 <BEV>

A:Cross-references: EMBL:AL035440

A:Experimental source: cultivar Columbia; BAC clone F10M23

C:Genetics:

A:Map position: 4

A:Intons: 395/1; 613/3; 777/1; 826/3; 1945/2; 1993/3; 2061/3; 2129/3; 2178/3; 2295/3; 2

A:Note: F10M23.350

Query Match 2.6%; Score 134.5; DB 2; Length 2535;

Best Local Similarity 19.9%; Pred. No. 1.5;

Matches 213; Conservative 128; Mismatches 336; Indels 391; Gaps 53;

Qy 21 DDFEFOAL---LDIMPEEKPLPTAFVDTFSEA-----LLPDLMLKLT--RMIRSEVRL 70

Db 568 EDAEKEYLGIVSDIWSELCSNP-----IDSVEEAEWCEFIKILKIDSLIKIYVRAVPNELGGS 623

Qy 71 VDAALDLEPQOLLFFVQSGIPVSSMKLQFLDQAVAHDPQTELEONIMDKNYAHLYE 130

Db 634 FDIFFKPLS-----NSFGMPVELQRLALLSLSEYISWTPKS----- 659

Qy 131 VOHERGASGGQTFHSLTLASLPPRRDSTEARPKSPSPQIGQRIYGTQRLVGLPED- 189

Db 660 -QSDSG-----PTRLP--PLMHKHLRVINILLSPHNG 690

Qy 190 --DLA-----GMFLQIF-----PLSPDPRMOSSPREVALA 218

Db 691 VKDLAVNLAVAMNSTGAFFENNPSEIGAMFLPLPCFEKIKLPLELOEAVQMSVAVSFL 750

Qy 219 LQ--GALGOELAR--VVOGS--PEVPGIYR-----VLQALATLLSS-----PHGA 259

Db 751 CDAVSTVGNNLFKHWIDIVRSSLSHKGVISGFSPPLITCLLQKCVRLNSESSTSLPEKSA 810

Qy 260 LVMSNHRSHFLACPLLRQLQYQRCVPODTGFSFLKVLQMLQMLDIPGVEGGLPRAQ 319

Db 811 ISL-----YVCSFLKYLQIQ---VDSKLISCLQSVLSEVDESKSLCEWRRLR-- 858

Qy 320 LRLMASQASAGRLSDVNGGLLR-----LAEALAFRODLEVASTVRAVIAITLNSGEOC 373

Db 859 MLTCSQS-----LSNERPIILHSRTTGLPADSSFAETLDEIKRLVR----- 901

Qy 374 SVEPRLISKVLOGGLEVNSPHLELLTAFFSATADAASFPRACKPVVVVSSLLQEEPL 433

Db 902 SISPEIRIGIYKAFSSALICATPESTILQFASVMDVSNAFYQ--TPPSFQSTITPLEENFL 960

Qy 434 AGKRGADGSGLEAVRLGPPSSGLVDWLEMLDPEVAVSCPDQLRLLEFRRKGQCAQVP 493

Db 961 GN-----LSKLSPLDLFASSE----- 976

Qy 494 SFRPVLTLFTHQSSWPTLHQCIRVLLGKSRQRPDPSSASLDLFLACIHPVPIWGGROR 553

Db 977 -----FTGS-----GNLCEGYVD--SEIDF-----SGHS-- 998

Qy 554 TPQKRREELVLRVQGPRLISIVELTLAEAFETPSOGDPTAACGLIQAURLLLSCCGDDE 613

Db 999 -----SYTEETRISKMDNDMESSAFSIFLKOAPPPVLLINAIMSMOI 1039

Qy 614 S-----VRKTEHLSCIOQMGDSVIGRRCRDLLLOLYLORBELRPVPEVLL 661

Db 1040 SCLPERRISELLLLKVSQPKSGSI-----DS-----NIQLTFWLFOIRSSYKQ--PAPVL 1090

Qy 662 HSEGAASSVCKIDGLIHRFTLLADTSDSRALENQAD---ASMCRLLAVAHPLLRL 718

Db 1091 HO-----LSEIC-----LRLMKNLFSQISEPELVSGPSSNKLPAFAFKHQAVERVLC-- 1139

Qy 719 HLPMTAALLH-----GTHLNFQEFQONHLSCLFHYGLL 754

Db 1140 H--PVVALLLESPLDCGTLPPVQVNEIFSETSLTMGR--LVFSEIDQ-----HILNLL 1188

Qy 755 ELQHPVFRSEHOGALM-----DCLLSF-----ITLLNLYR-----KSSRLH 791

Db 1189 VSTCEHFLFDEKRPNNMKRDKLKNKSIYAFKDLVEKLLLEFRKFLCVCQSOSYSLLQ 1248

Qy 792 AAFINKEVQFIHKYTYVNAAPAIISFLQKHADPLHDSFQNSDLVMSKL--LAG-----LS 845

Db 1249 AOLIHALLFISPFKLFN--IAHSMCLKIDE--EGLTSPNSIIISLGLIGAGAFEMLY 1304

Qy 846 LPSRODRTRGIDE-----EGEESASAGSL-----PLVSYLFTPLTAEMA----- 887

Db 1305 LVSHQFTAKRGVYDILMELEERNYASNIIEKYYSMAKCFSTSLDSDADICLKVCGGIT 1364

Qy 888 -----PYMKRLSR--GQTVEDLLEVLSDIDEMSRREPLISF 922

Db 1365 RCKHNQNVAVDPLVNLKISLVGRTPEDL--IHCIRASITRAKILFY 1410

RESULT 8

AE0119

CIP ATPase [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001

C:Accession: AE0119

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G

ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; M01D:21470413; PMID:11586360

A:Accession: AE0119

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-882 <KUN>

A:Cross-references: GB:AL590842; PIDN:CAC89816.1; PID:q15979042; GSPDB:GN00175

C:Genetics:

A:Gene: CIPB3

C:Superfamily: endopeptidase C1p ATP-binding chain

Query Match 2.5%; Score 131.5; DB 2; Length 882;

Best Local Similarity 21.3%; Pred. No. 0.5;

Matches 170; Conservative 95; Mismatches 279; Indels 253; Gaps 36;

Qy 22 DSEFOALDIMPPEEKPLP-----TAFVDTSEBALLY-----PMLKLRMR 64

Db 220 DSETRQMDILSRKRNKNLIVGEPGVKTYL-----EGLALRIEAGNVPSLKTQSVR 274

Qy 65 SEVLRVDA-----ALODLEPQOLLFFV-----OSFG----- 91

Db 275 TLDLGLQAGAGVKGFEFORLKNVLEAVQO--SPSPVLEFIDEAHTLIGAGNAGGADAAN 333

Qy 92 -----IPVSSSKLLOFL--QQAVAHPDQTELEONIMKMYAHLYEVOHG 136

Db 334 LKLPALAGELRTIATVTSKQYFERDAALERRFQMKVDEPDDTACMLRGLKERY 393

Qy 137 AS--GGQTFHSLTASLP--PRDSTEARPKSPSPQIGQ--RIR-----VGT 180

Db 394 ATHHGVHILDAITAAVTLSRRFLTGRQRLDRAVLDLDTAGARVRSIDTFLTALMAINA 453

Qy 181 QLRVLGPEDDLAGMFLQIFPLSPDPRMOSSPREVALAL--QALGOELARVVQGSPEVPG 239





## A:Map position: 1

Query Match 2.4%; Score 128; DB 2; Length 1021;

Best Local Similarity 21.7%; Pred. No. 1.1;

Matches 195; Conservative 74; Mismatches 314; Indels 314; Gaps 40;

QY 121 DKRYMAHLV-EVCHENGASGGQTFHSLTLASLPDRDSTAPKPKSPEOP---ICGR 175  
 DB 70 DKRLLELRELAQEAQL-----RDLGAPAEQEPAPLPPELELLGRSK 112  
 QY 176 IRVGTOLRVLPEDDLGAMFLQIFPLSPDRMQSSSPRPV-----ALAL 219  
 DB 113 KRYVKEP-----LAGV-----RAGRTPLVYVGTGAGAGKTVLLDHLAAL 152  
 QY 220 QQA-----IQGELARVVOGSP---EVPGITVRVLAALTLSS---PHGALVSM 264  
 DB 153 QGAGARVERLNLGDLALAGPPLRDRSPASLAARQDLALRYVLPERPVPVGSALLVRY 212  
 QY 265 HRS-HF-----LACPLRLQCOYORCVQODTGFSSLFKLVLQMLQW- 305  
 DB 213 TENLHFGNDPRLPDGTPATAAAMAEVLR-----RHIF--AGVSYLF--ALEDAAGWP 262  
 QY 306 -----LDSPGVEGGLRAQOLMLASQASAGRRL- 333  
 DB 263 PGAGELIELQPTLTERATILMARLVSRVADSLARETRNDRLTLGAGGNARL 322  
 QY 334 --SDVRGGLRLAEALAFRODLEVSSTVAVIATLRSQCSVEPDLISKVLGLEIEVR 391  
 DB 323 ADHVDV---RLACAAAV---LAVPPTGTDTASSVPLGDPVPLPAVALALALGGVYAL 375  
 QY 392 SPHLEELTAFFSATDASFPACKRVVYVSSLLQEEPLAGKRGAGGSLAVRLG 451  
 DB 376 PLARSILDM-----VTPPLEGEGPAGQAVGEGTG--EKYTGW 413  
 QY 452 PSSGLVDMLE-LDEEV--VSSCPDLQLRLFSRRKKGQAOVSEFPYLLFTTHQSSV 509  
 DB 414 RRPRLP--WALPLRPEAEVATEQVQATQSRDLS---PEVVPYLLALALAGEM 465  
 QY 510 PTHQCIRVLGKSGREOPDPSASLDFLWACIH-----VPRIMOGDRORT 554  
 DB 466 SMLEHVRRA-----RPDDARYLLPALMRIRIAGASSPEREGARAVVTHVASRGEYHA 517  
 QY 555 POKRRELYRVQGPPELISVELLAEETRSDODGTAAASLQARPLLLSCCGDES 614  
 DB 518 PAARDALFVLEEGAAHAARVLAESSLDAGNPFETAOTOLEKAEI----- 564  
 QY 615 VRKTEHLSCGCIQWGS-VLGRRCRDLILQLYLOR-----PEL 652  
 DB 565 -----EHLVSQPGSGWGDWRLAOSGLVQALARWRGMAGATEAVSPRTAQSFPRA 619  
 QY 653 R-----VVPYEVLLHSEGAASSVCKLIDGILHRTITLLADTSD 690  
 DB 620 RLWRGLVADKAGHWEALDHLGAVPGSSPLSTRARYQEGDRLR--RLGQPVYAYALSD 677  
 QY 691 -SRALENRGADAMACRKLAIVAPRLLLRLPRITALLHRTHLNPFQEQNHLSFLH 749  
 DB 678 AARRLAAGANPEQARVLA---ATARRRGYAE---DALRLFGQ 718  
 QY 750 VLGLLELLOPHVRSEHOGALMDCLS-FIRLLINRKSSRHLA----- 792  
 DB 719 ALALL-----PTDLRSADVLAQRLISEQVPLVLAGRPDALLQAQAALITLLRGEGGR 774  
 QY 793 APLNKKVVOFLH---KYTTYNAPAAISFLQKADPLHDLSFNSDNLVMLKSLAGL 844  
 DB 775 AEVTYEVRRTHYRVALAYLTRG--RGVPYLOPFGP---EFDTPDLVHARALLAEI 825

## RESULT 13

AE0179  
 Probable ATPase chain of ATP-dependent proteinase YPO1471 [imported] - Yersinia pestis  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AE0179

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.  
 demo-Tariga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MIMD:21470413; PMID:11586360

A:Accession: AE0179

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-880 &lt;KUR&gt;

A:Cross-references: GB:AL590842; PIDN:CMC90296.1; PID:q15979515; GSPDB:GN00175

C:Gene: YPO1471

C:Superfamily: endopeptidase C1p ATP-binding chain

Query Match 2.4%; Score 126; DB 2; Length 880;

Best Local Similarity 21.6%; Pred. No. 1.2; Mismatches 271; Indels 230; Gaps 34;

QY 22 DSEFOALLDIWFPEKPLP-----TAFVDTSEALL-----PDWLKLRMR 64  
 DB 220 DSEIRQIDILSRRRKNNPILVGEVGYKTALV-----EGALRLADGNVPDSKTVSVR 274  
 QY 65 SEVRLVDA-----ALQDLPEQQLLFV-----QSTG----- 91  
 DB 275 TLDTGLQAGAGVKEFEQRLKNVIEAVQO--SPSPVLLFIDEAHTTIGAGNAGADAN 333  
 QY 92 -----IPVSMKLLQFL--DOVAHDPTLCNIMDKVMHMLVCHVCHERG 136  
 DB 334 LKRALARGELRTAATWSEYKQYFERDAALERRPQWVYDEDDDTACMLRGLERY 393  
 QY 137 AS--GGQFHSLSLASLP--PRDSTEAPKPKSPEOPIGG--RIR-----VGT 180  
 DB 394 ATHGVVILDAALTAATLSRREFLQRLQPKAVDLDTGAGARMSIDTLPTALMEINA 453  
 QY 181 QLVRLGPEDDLQAGMFLQIFPLSPDRMQSSSPRVATL--QALGQELARVVOGSPVPG 239  
 DB 454 ELAALAMEQQAIEDDLLLPVVSSTRLEPTEQRAALAVGCHTLEQY-----KEEK 506  
 QY 240 ITVRVLOALATLLSPHGALVMSMRSHFLACPLRLQCOYORCVQODTGFSSLFKVL 299  
 DB 507 LTTILIEARQDINAAH-----LVALQELVQIQGNAP-----LSTLDVDR 548  
 QY 300 LQMLQWLDSPGVEGCPRLAOLMLASQASAGRILSDVRGGLRLAEALFR--QD--LE 354  
 DB 549 TVAAVIVDMWGV--PL-----GSLKDEQTMLSLENRLGERVIEQDAALG 592  
 QY 355 VVSTVAVIATLRS-----GEOCSVEPDLISKVLQGLIEVRSPHLELLTAFPSATADA 409  
 DB 593 ALAQRLEAAKTGLISENGPLGVLVLPVSGTGT-----ETALALA 633  
 QY 410 ASPFPACKPVVNVSSLLQEE--EPLAGKPG---ADGSL--EAVRLGSSGLVDMWL 461  
 DB 634 DSLFGKSKSLITINLSYEQAHTVSOQLKSPGVGVGGGVLEAVRKRYSVLLDEV 693  
 QY 462 EMLDPEVSSCPDLQRLFSR--RKQGAQOVPSR-----PYLLTFTHQ 506  
 DB 694 EKAHRDVLN-----LFQYVDRDGRFMDEGRE--IDFRNVIVYWTANLGSCHLMQLDEQ 746  
 QY 507 SSMP--TLHCIRVLGKSGREOPDPSASLDFLWACIHVRIMWGRGQRTPOKREELV 564  
 DB 747 PDADOSTLHELRLPIL---RDHFQPLARF-----QTLTY 779  
 QY 565 R-VQGPPELISVELLAEATRSDODGTAAASLQARPLLLSCCGDDSVKRVTEHLS 623  
 DB 780 RPLQYDALKRITVAIKLNQVAGNRHLCHGSCQIEDSLDTLVAACLLPDVGARNI----- 834  
 QY 624 GCTQMGDSVYGRRCRDLILQLYLOR 649  
 DB 835 -----DSLNLQOILPVLSQLSLR 853

## RESULT 14

113057  
 KIAA0729 protein - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13057; T08741  
 R:Nagase, T.; Ishikawa, K.; Suyama, M.; Kikuno, R.; Miyajima, N.; Tanaka, A.; Kotani, H.  
 DNA Res. 5, 277-286, 1998  
 A:Title: Prediction of the coding sequences of unidentified human genes. XI. The complet  
 A:Reference number: Z17595; MUID:99087487; PMID:9872452  
 A:Accession: T13057  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1196 <NAG>  
 A:Cross-references: EMBL:AB018272; NID:93882178; PIDN:BA34449.1; PID:93882179  
 A:Experimental source: brain  
 R:Mambuti, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, March 1999  
 A:Reference number: Z16471  
 A:Accession: T08741  
 A:Molecule type: mRNA  
 A:Residues: 279-1196 <NAG>  
 A:Cross-references: EMBL:AL050092  
 A:Experimental source: adult uterus; clone DKFZp586G0518  
 C:Genetics:  
 A:Note: KIAA0729; DKFZp586G0518.1  
 Query Match 2.4%; Score 125.5; DB 2; Length 1196;  
 Best Local Similarity 20.0%; Pred. No. 2.1; Indels 241; Gaps 39;  
 Matches 155; Conservative 101; Mismatches 277;  
 QY 213 RPVALALQALGQELARVVOGSPEVP-----GITVRVLAQALATLISPHGALVMSMR 266  
 DB 35 RPVAHNLVLPQME-----DGSDDMDTSVEDIGRSCTVRFRTLLIMEHGVKPSKRL 89  
 QY 267 SHFLACPLRLQCYQRCVPDPTGSSFLKVLQMLQMLDSPGVGGELRLQRLMASQ 326  
 DB 90 TEYFAF-----LYEPKMGKEESQF-LLSLQALISTWVHFY--MGTK-GPENQVEVLSEE 140  
 QY 327 ASAGRLSDVRCGLRLAEALAFRODLEVSVTRAVITLRSGECSVE----- 376  
 DB 141 EGEEEEE-----DLSLAEKTRPALEKMLALVLYOOSERILLISQDMAALTG 197  
 QY 377 ---PDLISKVLQGLIEVR-----SPHEELL-TAFESATADASPPAC 416  
 DB 198 KGPEPLFQHIRDG-INIROTCNLIFSLCKRYNNRLAEHIVSMFTSTAKLTPEANPE--- 253  
 QY 417 KPVVVSSLLQEEERPLAGKRGADGS-----LEAVRLPSSGLVDMLEMDPE- 467  
 DB 254 ---FKLLTLMLE---FAGGPRGMPFPASYILQRIWEVLEYNPSOCL--DWLAVOTPRNK 304  
 QY 468 ---VSSCPDLQRLLEFSRRKGGQAVPSFRPYLLTLFTHOSW-----P 510  
 DB 305 LAHSVNLQNMENWVERFLAH---NYPRVTSAAVLLVSLIPNSFRQMFSTRSLHP 360  
 QY 511 T-----LHQCIRVILG-KSRQRFPSA-----SLDFLMAC----- 540  
 DB 361 TRDPLSPDTVVLHQVYVNLGLLSRAKLYDAVHGTTKLVAFPSFTYCLISTEKL 420  
 QY 541 -----IHPRIINGRQRTPKRR-----EELVLRVQGPBELISLVE 576  
 DB 421 MESTYFMDLMLNFPKLESPATATNHNKQALLSFYVNCADCPENIRLLVQNPVYTKNTA 480  
 QY 577 LILAEETRSQDGTAAACSLIOARLPLLLSCCGGDESVRYKTEHLSGCTIOQMGDSVLK 636  
 DB 461 FNYLIADDDQDVLFNGKMLPAYYGILRLCCGSPATFRLASHON---IOWAKNLTPE 537  
 QY 637 RCRLD-----LLQLYL-QRPELRYVPEVLLHSEGAASSVCK-LDG-----LTH 679  
 DB 538 HASQYPAVEELFNMQJFIACQRPDMREBELDIKQFKTTISCYLRCDGSCWTTLS 597  
 QY 660 RFTTLAOTSRALE--NRGADASMACKLAVAHPRLLLRLPMTAALLHGRTHINPE 737  
 DB 598 AF-RILLESDEDRLLVFNRG-----LILMTE-----SFNT 627

QY 738 FROQNLSCFLHVLG-LLELL-----OPHVERSEHOGAL--WDCILSFIRELLNY 784  
 DB 628 LHMHTAEATACHVTGDLVELLSFLSVLAKSTRYLQKDYKQALIONQERIEFAHLLTL 687  
 QY 785 RNS-----SRHLAAEFINKVOFT-HKITYT--NAPAI 814  
 DB 688 LMSYSPPELRNACIDVLEKELVLSPP--DELFTLVPELQHNCTYHNSNPML 739  
 RESULT 15  
 C83110  
 Probable exonuclease PA4282 (imported) - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 14-Sep-2001  
 C:Accession: C83110  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
 Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kass, A.; Lapidig, K.; I  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: AB2950; MUID:20437337; PMID:10984043  
 A:Accession: C83110  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1211 <STO>  
 A:Cross-references: GB:AE004844; GB:AE004091; NID:99950500; PIDN:AA07670.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4282  
 C:Superfamily: sbcc protein  
 Query Match 2.4%; Score 125.5; DB 2; Length 1211;  
 Best Local Similarity 21.0%; Pred. No. 2.1; Indels 426; Gaps 49;  
 Matches 232; Conservative 118; Mismatches 426;  
 QY 24 EFCALIDI-----WPEERPLPAFLVDTSEBALPDMILKIMIRSEVLRYDALODL 78  
 DB 251 EQALQRLRGQOQWTFEQR-----LLQSEHA-----QGLAFAKQAADALATER 226  
 QY 79 EPOQLLEFVQSGITPVSSSKLLQFLDQAVADPQLEQNTIMDKNMAHLYEVQHERGAS 138  
 DB 297 ETLQWLERL---APVGLIERLQLELRHSEQOQOQ-----RTQQA 338  
 QY 139 GGQTFSLTLASLPFRDSTEARPKSPQPRGGGRIRVQTL-----RVLPEPD 190  
 DB 339 GTERLQGLARLOEAREQAOADNHLROAOP-----LEAPOLESEARLENTLAEROE 393  
 QY 191 LAGMFLQIFPLSPDPRWQSSSPRPVALALQALGQELARVVOGSPEVPGITVRLQALAT 250  
 DB 394 L-----HRQSNORHAQOSDAARQMDMQQRHVARQALQA-ALRDSQALAA 438  
 QY 251 LLSS--PRGALVMSHNRSHFLACPLRLQCYQRCVPDPTGSSFLKVL-----LQML 303  
 DB 439 LQDAWVTHQGLATFVQRR-----QRLAESQALPBEKSLAHAGEPLERL 484  
 QY 304 ---QWLDSPVEGGRPLAOLRMLASQASAGR-----LSDVAGGLRLAEALAFRQ 351  
 DB 485 QAWTALHSGSEPDLAARLVELRLQTDLSERQALHKEMQQVLDQAGLARIGEL----- 540  
 QY 352 DLEVSSIVRAYATLRSGEQCSVEPDLISKVLQGLIEV-----RSHLEELLTAFFS 404  
 DB 541 DQRMVQGE-QALLDLKRGSSQCAEEYKAAQALQVRELLQRLRARSASVEQLRGLVD 599  
 QY 405 ATADAASPPACKPVVVVSSLLQEEERPLAGKRGADGSLAVRLGPSSGLLVDMLEML 464  
 DB 600 GEA-----CPVCG-----SCQHP-----YHNEQLLAALGEHD 627  
 QY 465 DPEVSSCPDLO-----LRLIFS---RRKGGQAVPSFRPYLLTLFTHOSWPTLS 513  
 DB 628 DQEQVAEBSLELRQTLVGLRGYSYSORERLNOSQEOBELTQGLAALDRQIDQW-TLP 686  
 QY 514 OCIRVLLGKSRQRFDPDSASLDFLMACIHVPRIWGRDQRTF-QKRREELVLRVQGPBELI 572

```
Db 687 EELRL-----OPSADLEWL-----AQRDLLAGQRQCCQRPDRLLARQRTQOL 732
QY 573 SLVELLAEAEFRSQDDGTACSLIQARLPLLSCCGDDESVRKV--TEHLSGCIOW- 629
Db 733 ---QOELEAEFTILQQRQALTEQRQ- RYEHLOQOVEDSQOLRPLLSDEHW-----QRWQ 784
QY 630 -----GDSVLRGRCDLLQLLYLQR- PELRVVPE-----VLHSEG 665
Db 785 ADPLRTFQALGESIEGRROOARLQIEQRLQELKQRCDDESSWOLKQSDQCRNEARQAE 844
QY 666 AASSVCKLGLHRTFTLLADTSD- SRALENRGADASMCRKLAVAHPILLRHPMA 724
Db 845 RAQELAEELNGRLGAHLGQHCAQDWQLSLEHAAQAQSAVETLQA--PLDSLREQL- 900
QY 725 ALLHGRTHLNFQEFROQNHU-----SCFLHVIGL-----L 754
Db 901 RLAEALHLLQOQRQDEFORLQADWQAMREODNLDSDRLDALLGISEQATQWRQL 960
QY 755 ELLOPHVFR-----SEHQALWDCLLSFIR-----LLNRYKSSRHIAFINKFVQ 800
Db 961 QRLOEETIRQOTLEAEHQAO---LLOHRRQRPETDREALEDNLRQQRERLASEQAYLE 1016
QY 801 FHKYITYNAPATSFLOKHADPLHDISPDNSDLVMLKSLAGLSLPSRODRTDGLDEE 860
Db 1017 -----TYS-----QLOADNQRQESQALAELEERARAEFRWRGRLNE- 1053
QY 861 GEEESSAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQTVEDLEVLSDID--EMSR-- 915
Db 1054 -----LIGSSGDKFRRIAQGYNL-DLIVQHSNVQLRLARRYRL 1092
QY 916 -----RPELISFSTNLQRLMSAEECCRNLAFLSLRSMQNSP-STAA 958
Db 1093 QRGSELGLLVDTMGMDELRSVYS-----LSGGETFLISLALGLASMASSKLRIES 1146
QY 959 AFLPTFMVYCGSDQDEEVQOTALRNL 983
Db 1147 LFIDEGFGLDPESLQLAMDALDNL 1171
```

Search completed: May 7, 2003, 19:42:51  
Job time : 42 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:38:42 ; Search time 15 Seconds

(without alignments) updates/sec  
2845.276 Million cell

Title: US-09-929-769-7  
Perfect score: 5241

Sequence: 1 MHILVHAMVILTLGPPRA.....MDPSAQISEALRIILMEAVM 1029

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	2.8	1302	1 ACSC_ACEXY	P37718 acetobacter
2	147.5	2.8	878	1 RPO1BDVA	P12918 avian infec
3	138	2.6	2248	1 Y539_HUMAN	O60287 homo sapien
4	138	2.6	3674	1 SPCR_HUMAN	O90268 homo sapien
5	133	2.5	1083	1 T2D3_HUMAN	O00168 homo sapien
6	127	2.4	3321	1 PCN2_HUMAN	O95613 homo sapien
7	125.5	2.4	1211	1 SBCC_PSEAE	O91468 pseudomonas
8	123.5	2.4	1455	1 FACA_HUMAN	O15360 homo sapien
9	123.5	2.4	2564	1 SPCO_HUMAN	O91254 homo sapien
10	123	2.3	817	1 MURS_THERC	O94166 thermus cal
11	122.5	2.3	738	1 COG2_HUMAN	O14746 homo sapien
12	122.5	2.3	2035	1 EYPL_MOUSE	O96852 mus musculu
13	122	2.3	1013	1 PTX_MACNE	O02895 macaca neme
14	121.5	2.3	1058	1 GEM4_HUMAN	P57678 homo sapien
15	121	2.3	1065	1 VINC_CHICK	P12003 gallus gall
16	121	2.3	1133	1 VINC_HUMAN	P18206 homo sapien
17	120	2.3	3066	1 ATM_MOUSE	O62388 mus musculu
18	120	2.3	3149	1 TEGU_EBYV	P03186 epstein-bar
19	119.5	2.3	818	1 MURS_THERH	O56339 thermus the
20	119.5	2.3	1115	1 TRC2_CHLRE	O8V9P3 chlamydomon
21	119.5	2.3	2541	1 TAL1_MOUSE	O94190 homo sapien
22	118.5	2.3	1237	1 B3A2_MOUSE	P13608 mus musculu
23	118.5	2.3	1237	1 B3A2_RABIT	P48746 oryctolagus
24	117.5	2.2	1065	1 SPCP_RAT	O94988 rattus norv
25	117	2.2	1065	1 VINC_MOUSE	O64727 mus musculu
26	116.5	2.2	804	1 RSG4_HUMAN	O95924 homo sapien
27	116.5	2.2	1234	1 B3A2_RAT	P23347 rattus norv
28	116.5	2.2	2541	1 TAL1_MOUSE	P26039 mus musculu
29	116	2.2	465	1 VP19_HSV1	P32888 herpes simp
30	116	2.2	5038	1 RYR1_HSV1	P21817 homo sapien
31	115.5	2.2	2390	1 SPCP_HUMAN	O15020 homo sapien
32	115.5	2.2	3746	1 ACVS_PENCH	P19787 penicillium
33	115.5	2.2	3791	1 ACVT_PENCH	P26046 penicillium

## ALIGNMENTS

RESULT 1	ID	ACSC_ACEXY	STANDARD:	PRT: 1302 AA.	
AC	P37718:	ACSC_ACEXY			
DT	01-OCT-1994 (Rel. 30, Created)				
DT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Cellulose synthase operon protein C precursor.				
GN	ACSC.				
OS	Acetobacter xylinus.				
OC	Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;				
OX	Glucanacetobacter.				
NCBI_TaxID=28448;					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 53582;				
NCBI_PubMed=8083166;					
RA	Saxena I.M., Kudlicka K., Okuda K., Brown R.M. Jr.:				
RT	"Characterization of genes in the cellulose-synthesizing operon (acs				
RT	operon) of Acetobacter xylinum: implications for cellulose				
RT	crystallization.";				
RL	J. Bacteriol. 176:5735-5752(1994).				
RP	TOPOLGY.				
RX	PubMed-11544230;				
RA	Kimura S., Chen H.P., Saxena I.M., Brown R.M. Jr., Itoh T.:				
RT	"Localization of c-di-GMP-binding protein with the linear terminal				
RT	complexes of Acetobacter xylinum ";				
RL	J. Bacteriol. 183:5668-5674(2001).				
CC	- FUNCTION: Required for maximal bacterial cellulose synthesis. It				
CC	may be involved in the formation of a membrane complex for				
CC	extrusion of the cellulose product.				
CC	- PATHWAY: Bacterial cellulose biosynthesis.				
CC	- SUBCELLULAR LOCATION: Outer membrane (Potential).				
CC	- SIMILARITY: BELONGS TO THE ACSC/BSC FAMILY.				
CC	- SIMILARITY: CONTAINS 7 TPR REPEATS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL: X54676; CAA38489.1;				
DR	InterPro: IPR003921; CellSynth_C.				
DR	InterPro: IPR001440; TPR.				
DR	Pfam: PF00515; TPR, 4.				
KW	PRINTS: PRO1441; CELLSYNTHASEC.				
KW	Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane;				
KW	Signal.				
FT	SIGNAL	1	32	POTENTIAL.	
FT	CHAIN	33	1302	CELLULOSE SYNTHASE OPERON PROTEIN C.	
FT	REPEAT	64	97	TPR 1.	
FT	REPEAT	147	180	TPR 2.	

FT REPEAT 306 339 TPR 3.  
 FT REPEAT 340 373 TPR 4.  
 FT REPEAT 572 605 TPR 5.  
 FT REPEAT 716 749 TPR 6.  
 FT REPEAT 750 783 TPR 7.  
 SQ SEQUENCE 1302 AA: 138750 MM: CC1E32F0E1F9D794 CRC64:

Query Match 2.8%; Score 148; DB 1; Length 1302;  
 Best Local Similarity 20.8%; Pred. No. 0.043;  
 Matches 194; Conservative 120; Mismatches 351; Indels 268; Gaps 43;

QY 102 QPFDQAVHDPQTEQIMKNMAHL-----VEVOHE-RGASGCTFFSLTASLPPR 154  
 DB 85 QSLQSAARLAPDVLLEAGEYQSHGNDAIDTORLHQAPAGSTYSQDLDL--- 141  
 QY 155 RDSLEAPKPKSPPEQIGRIRGQIRVQLRVLGPEDDLAGEMLQIFPLSPDPMQSSPPR 214  
 DB 142 -----HEQATISQDILAHARSLASGHSQDAVEAYQHLE-----NGSTFTP 181  
 QY 215 -VALALQAL-----GQE-----LARVQGSPEYVGTIVYVLAQALATLSSPHIGALVMS 263  
 DB 182 SLAVEYVQTLGAVSGQAGTAGDGLIRLVKANPS---DFAQLALAQVLTLY-QPCTRMEG 236  
 QY 264 MHSHFLACLLRLOLOYOYCPQDIFGSSLFLKVLQMLQMD-----SPGVEGG 314  
 DB 237 LQR-----LQALQATQSSAPAEATAE---KSYROTLSMLPVTPETPLMOKMDAH 285  
 QY 315 PLRAQLMLASQASAG-----RLSDVRGGLRLAEALAFRODELVSSSTVRAVI 364  
 DB 286 PDSALRTNHAEPAGGPPDGALARODGFALNAGRISAQA-AFOSALMLNK----- 338  
 QY 365 ATLRSGQGVPEPDLISKVLQGLEIVNSPHLE---LTAFFSATADAASPPACKPVV 420  
 DB 339 -----DDDALGL--GLVAMRAGHNEARLYLEDAIADPPKNAHWMPRA----- 380  
 QY 421 VVSLSLQEEPLAGCKRPGADGSL-----AVRLGSSGLVYDML 462  
 DB 381 -----LAGMAVGEISGVNRLLASGTOEAQRLMTLARPQSGEATL----- 424  
 QY 463 MLDPEVVSQPDQLRLLFSSRRKQ-GQAQVPSFRPLTLFTTHOSSMPTLHQCIRVLG 521  
 DB 425 ML-----ADLQ-----RSTGQTEAE---RNRRLALRANGDPIALMGLARVLMG 466  
 QY 522 KSPQRPDPASIDFLMACIHVPRIWQDQRTPOKRREELVAVQGPBLSIVELLAE 581  
 DB 467 EGDES--EANALLSRLOG-----RYSQVOQIE-----VSGIMAE 499  
 QY 582 AFRSODQDTRACSLQA--RLPILLSCCGDESVKRVTEHLSGCTIQMGDSVLRRCR 639  
 DB 500 A-ARTSDSAOKVSLRQAMTKAP-----DDPMLR---INLANALQOOGDSSEAN-- 545  
 QY 640 DLLLQVLYLRPELRVPVPEVLHSEGAASSVCKLDGLIHRFTLLADTSDSRAL---E 695  
 DB 546 --VMRPLLTSPRTPADYQAAILVASGNGN-----DTLARLLAGLSPPDYSPAIRTIAD 597  
 QY 696 NRGADASMACRKLAVAHPILLRLPMTIALHLGRTHLNQEPRQNHLSGFLHVLGLE 755  
 DB 598 EMAIKADLASRLSMVSPPTLVREALAAPDTPGARVAVADLFRQK-----D 645  
 QY 756 LLDPHVFRSEHOGALMCLSTFIRLLNLYRKSRRHLAALFNKEVQFHKYITYNAPAI 815  
 DB 646 MLRAHM-----ALRI-----ASTRNIDLTTEQRLAVYTEVTKISNPYAAA 685  
 QY 816 FLQKHADPLHDSFNSDLVY-----IKSLLAGLSLPSRDORTDGLDEGESESSAG 868  
 DB 686 RL---LAPLDGSGTAGTASMSPDQROTLMQIRNGISVAGSDLLNGG-DQAAAYDHLAP 741  
 QY 869 SLPLVVSLETPPLTAEMAYMKRLSGQVEDELLEVLSDIDERSRRPELTSFSTNLQ 928  
 DB 742 ALQADPEATSPKALALAL--YNGRGKYGHALDIDLAVL-----RNPDDLARQAAVQ 792  
 QY 929 RLMSAEECCRNLAFLSLRSMONSPSIAAFL 961

DB 793 AAANDKND---NLAMQALQDGVQOOSPMDARSWL 822

# RESULT 2

RRPO\_IBDVA  
 ID RRPO\_IBDVA STANDARD; PRT; 878 AA.  
 AC P12918;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative RNA-directed RNA polymerase (EC 2.7.7.48) (VPI protein)  
 DE (RDRP).  
 GN VPI.  
 OS Avian infectious bursal disease virus (strain Australian 002-73)  
 OS (IBDV).  
 OC Viruses; dsRNA viruses; Birnaviridae; Avibirnavirus.  
 OX NCBI\_TaxID=10997;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88160058; PubMed=2831661;  
 RA Morgan M.M., Macreadie I.G., Harley V.R., Hudson P.J., Azad A.A.;  
 RT "Sequence of the small double-stranded RNA genomic segment of  
 RT infectious bursal disease virus and its deduced 90-kDa product."  
 RL Virology 163:240-242(1988).  
 CC - FUNCTION: THIS IS THE PRESUMPTIVE VIRION-ASSOCIATED RNA-DEPENDENT  
 CC RNA POLYMERASE. IT MAY ALSO CONTAIN GUANYLYL AND METHYL  
 CC TRANSFERASE ACTIVITIES.  
 CC - CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC (RNA)(N).  
 CC - PTM: MAY EXIST IN MULTIPLE PHOSPHORYLATED FORMS.  
 CC - SIMILARITY: BETWEEN THE CENTRAL REGIONS OF IPNV AND IBDV VPI.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M19336; AAA89177.1; -  
 CC PIR: A28649; RRSXB.  
 DR Transferase; RNA-directed RNA polymerase; Phosphorylation;  
 KW GTP-binding.  
 KW  
 FT DOMAIN 411 422 RDRP CONSENSUS ELEMENT-1.  
 FT DOMAIN 483 498 RDRP CONSENSUS ELEMENT-2.  
 FT DOMAIN 527 540 RDRP CONSENSUS ELEMENT-3.  
 FT NP\_BIND 258 265 GTP (POTENTIAL).  
 FT SEQUENCE 878 AA: 97274 MM: 6B67C1E0AC193D1F CRC64;  
 SQ

Query Match 2.8%; Score 147.5; DB 1; Length 878;  
 Best Local Similarity 20.2%; Pred. No. 0.026;  
 Matches 157; Conservative 109; Mismatches 272; Indels 241; Gaps 39;

QY 17 PPRADSEFQALDIPPEKPLPT-----AFIVDTSEAL-----LLPD 56  
 DB 22 PTAGQDVEELLTPKVVPPEDPLASPSRLAKFLRNGKYLQPSLDPENBEYEDQILPD 81  
 QY 57 WLKLMIRSEVL-----RLVDAALQDLEPOQ-----LLVQSF 90  
 DB 82 LAMWRQTEGAVLKRTLSLHMGPRVLPVLLNSPPEQKACQVPRHCTQADIVLFQ-- 139  
 QY 91 GIPVSSMSKLLQFDQAVHDPQTEQIMKNMAHL-----MAHLYEVOHERASG 139  
 DB 140 -VEPAT-----ESLQKEVTLTNTIRDKAAGSGTYGQATRLVAMEVATGRPNK 189  
 QY 140 -----GQTFHS---LTLASLPDRDSEAPKPKSPPEQIGGGRIRVGTOLRVLPEDDL 191  
 DB 190 DPLKLGTFESIAQLDITLP-----VGPEEDDKPVVPL---TRVPSRLVLY--TGDV 238  
 QY 192 AGMFLQIFPLSPDPMQSSPPR-VALALQALQGOELARVAVQGSPEVQGTIVYVLAQALAT 250  
 DB 239 DGDPE-EVEDLPLKINLKSSGGLPVYVGTGKETIGEMIA-----ISNQLRELSA 286



```

OY 251 LLSPPHG-----GALVMSHRSFLACPLRLQCO-YOR-----CVPQDTGFS 292
DB 287 LKQGGTGSNNKKLLMSLDQWYLSGILLFKAERYDKSTVLTTRNTWSPSPF--- 343
OY 293 SLFLKVLQMLQW---LDSP---GVEGGPLRAQLMLASQASAGRRLLDVGGLRLAE 345
DB 344 ---HLMISMITWPMVMSNSPNVNLIEGCP-----SLYKFNPFRCGLNATIVE 386
OY 346 -----ALAFRODLEVSSVRAVIATLRSGECCS---VEPDLSKYLQGLIEVRS 393
DB 387 WILAPEEPKALYADNIYVHSMTWYSIDLEKEANCTROHMOAAYITLITRQWSDNGDP 446
OY 394 HLEELTAFFSATADASPFPACKPVYVSSILQEEEPPLAGCPGADGSLFAVRLGPS 453
DB 447 MFMQWTASAFAMNIAFA-----LVYDSSCLIMNLQIKSYCO---GSGNAAFINNHLL 495
OY 454 SGLVDMLEML-----DPEVYSSCPD-LQRLFLFSRR---KKGQAQVPSFRPYLTTF 503
DB 496 STVLDDQWMLMKOPNDSEEFKSIEDKLGINFKIERSIDDIRGKLRQLVPLAQGYLS-- 553
OY 504 THQSSMPTLHQCIRVLGKSREORFDPASIDFL-WAC-----THVPRINGROQRT 555
DB 554 -----GGVEPQSSPVVELDLGWSATYSKDLGIYVPL----- 587
OY 556 OKRREELVLRVQGPFLISVELILAEETRSDGDTAACSLIO---ARL-----PLLLS 606
DB 588 --DKERLFCGSAAYPKGVENKSL-----KSKYGIQDAYVYVYEAIRLVGGMNYPILNK 638
OY 607 CCCGDESVRYKYTE---HLSCGICQMGD-SVIGRRCRDLLQLYLQY---PELRVVP 657
DB 639 ACKNNASARHLEAKGFLDEFLAEMSELSEFGEFEGFNIRKLTVREMLAEINKRVP 697

RESULT 3
ID 1539_HUMAN STANDARD: PRT; 2248 AA.
AC 060287; Q9NYQ01;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein KIAA0539 (Fragment).
GN C21ORF108 OR KIAA0539.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20237674; PubMed=10773462;
RA Slavov D., Hatori M., Sakaki Y., Rosenthal A., Shimizu N.,
RA Minoshima S., Kudoh J., Yaspo M.-L., Ramser J., Reinhardt R.,
RA Reimer C., Clancy K., Rynditch A., Gardiner K.;
RT "Criteria for gene identification and features of genome organization:
RT analysis of 6.5 Mb of DNA sequence from human chromosome 21.";
RT Gene 247:215-232(2000).
RL [2]
RN SEQUENCE OF 1427-2248 FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:331-39(1998).
CC -1- SIMILARITY: SOME, TO C. ELEGANS T05H4.10.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: AF231919; AAF22943.1; -
DR EMBL: AB011111; BAA25465.1; -
DR Genew; HGNC:117344; C21orf108.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 2248 AA; 252316 MW; AC0001821C002D61 CRC64;

Query Match 2.6%; Score 138; DB 1; Length 2248;
Best Local Similarity 19.5%; Pred. No. 0.45;
Matches 250; Conservative 185; Mismatches 428; Indels 420; Gaps 62;

OY 23 SEFOALDIFPEE-----KLPFAFLVDSSEELLPDMLKLMINSEVRL- 70
DB 830 SRFNRYSLMTIPQAREAWLLQAGSSPALPLASSFTLLQAAV-ESQALREHIOYO 888
OY 71 VDAALODLEPOQLL-----FVQSEFG-IPVSSMSKLIQ-FRD---QAVADHPQT 114
DB 889 LQATMPLMSQOYVLLAAKQYLVLRSTVENFGQGRVSGPPLQLFLDLRLVHCEQL 948
OY 115 LEON-----INDKNTMAHLEVQHERGASGCTFHSULTA----- 149
DB 949 DAONOQCEAARAADLDFLMESVAST-ELAND-----QTLREVALIRHPTLEGWFL 1001
OY 150 -----SLPRRDSPEAKF-----PKSSP-EQPIGO-----GRIR 177
DB 1002 ALBEOQALPPTLSLVYKLLATHTFSAGVLOLLASAVITLQNGILGARYSEAITQSVL 1061
OY 178 VGTQLRVLPEDDLAGNFLOIFPLSP-DPRWQSSPPVALAL-----QQ----- 221
DB 1062 KELQNRRAQPATSPKTPPQLEALQELHPYMEGAQLREVTLLALSLPETHLVYQOQPKSP 1121
OY 222 -----ALGOELARVQGSPE-----VPGITVRVLAALATLSSPHGAL 260
DB 1122 GKERNHLMAKTTLVOLLTCSPDOLQSGELMSSEYVRGA-----LLPTLADELDTVL 1177
OY 261 VMSMHRSHFLACPLRLQCOYORCVPODT-----GSSSLFLKVLQMLQWDSPEVEGCP- 315
DB 1178 LHTLQRPVLAIRAVGADLLDY---CLARTQALSTAILLIQESCTHLMFEQKCLQAPG 1235
OY 316 -----LRAQLRMLASQASA-----GRRLL-----SDVR 337
DB 1236 LGLOGLDDFLPLIHYVLQCRTHSHFRPAGVSAVIVPLRKTLMLRQLSLTSDSPA 1295
OY 338 GGLLR--LAELALF--KQDLEVVSVYRAVIATLRSGECCSVEPDLISKYLQGLIEVRS 393
DB 1296 SGLYQETLQVLPPARAKDLVMDRLPSLHTPPSSHKRWIV-ADSTISALEGSAEELCA 1354
OY 394 HLEELTAFFSATADASPFPACKPVYVSSILQEEEPPLAGCPGADGSLFAVRLGPS 453
DB 1355 WRKTL-----ESCYKWLIVS-----FSGGQDDNDTQNDQEKMLLR 1391
OY 454 SGLVDMLEMLDPEVYSSCPDLQRLFLFSRRKKGQAQVPSFRPYLT---LETHQSSW 509
DB 1392 LNALHMLNEVDP-----QDWQ---KFVKKGLKFRYQDHTFLKMLLTAVOLLTSPSSV 1442
OY 510 PTLHQCIRVLGKSREORFDPASIDFLMACINHPRIWQGRDQTPQKREELVLRQGP 569
DB 1443 RT--KLQL-----PVYVYMLQHSFLPTLLTSDGESPSQVKEALVLDL-- 1486
OY 570 ELISVELILAEETR-----QDQDTAACSLIQAIRLPLLSCCGGDESVRYKTEHNSGC 625
DB 1487 -MLTVEMCPVCESSHFAVLGAYGATLSVLDKILLIRARQNKLSL-----INFR 1539
OY 626 IQQMGDSV-----LGRRC-----RDLILQLYLQPELRVVP---E 658
DB 1540 VLLMGPAVAVHNHTCRSLGSLQWQPSVGDILRLDLDRMMQGITLHPQRRLLRPBDQ 1599
OY 659 VLLHSEGAASSVCKLDGLIHRTTLADTSDSALENRQADASMCRKALVAHPL----- 714
DB 1600 ELIFKD-----KSRVDDGLDVPCLFLQLFSELTR-----PEFVDCRKFLDSNALGLTV 1649

```



QY 173 -GGRIRVGTOLRVLPEDDLAGMELQIFLSPDRMSSSPRYALAQ----- 221  
 DB 3224 LQGMOKETAL--MKGED--GGHSL-----SSVR-----TLQOQRRRLERELE 3262  
 QY 222 ALGQELARV-----QGSPEVPGITVRYLQALATL--ISSPHGALVMSHNSHFLA- 271  
 DB 3263 AMEREVARLQTEACRLQGLHFAAPGGLAKVQEMATLQAKAQRGQWMLAQQAQGHAFGR 3322  
 QY 272 CPLL-----ROLQYQRCVQPDGFFSLFLKVLQMLQMDSCVEGCEPLRAQRLMALS 325  
 DB 3223 COELLANAQERQELASSEELAEVDVGAQQLQGHQELQDETRCRLQODLRQEQOQLVD 3382  
 QY 326 -----QASAGRRLSDVNGGLRLAELAFROD-----LEVSSTVAVIATLTSGDQC 373  
 DB 3383 NSHFNMSAEVTECQLQELDEGRQLQELAEVAMLRWQRCAESGCLKINGRLQGLQAEVLAQWGL 3442  
 QY 374 SVEPDL---ISKVLQGLIEVSPHLELLTAFFSATADASPFPACKPVVVVSSLLQEE 430  
 DB 3443 LKRDYGHVSVDV--ELLHRHODEKLLA-----QEEKFQOMQKTEMEQOELLLOPQ 3493  
 QY 431 EPLAGKRGADGSLAEVRLGPS--SGLLVDMLEMLDEPVVSSCPDLQLRLFSRKGK 488  
 DB 3494 EL-----KPRAGSSLTSPQMPRSGHQIGLAQLAETRPQDAKKGPTMGSLFEPKQHLPLG 3549  
 QY 489 QAQVPSFRPYLLLTFTQSSWPTLHQCIRVLLGKSREQRFPDPSASLDPLMACIHVPRWQ 548  
 DB 3550 GRQ-PS-----SSWMSD--CGTLQSSLSLFLDERMAAEKV-ASIALDLDTG 3593  
 QY 549 GRDORTQK--RRELVLRV--QGPPELLSVLELLAEATRSOD-----GDTAACSL---I 597  
 DB 3594 ARCLRLGRHGRKHTFSLRLTSGAEI-----LFAAPSEQAEASWRAUGSTAQSLSPKL 3648  
 QY 598 QARPLLLSCCCGDD 612  
 DB 3649 KAKPVSLNECTTKD 3663  
 RESULT 5  
 ID T2D3 HUMAN STANDARD; PRT: 1083 AA.  
 AC 000268; Q99721; Q9BX42; Q9BR40;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Transcription initiation factor TFIIID 135 kDa subunit (TAIFI-135)  
 DE (TAIFI135) (TAIFI-130) (TAIFI130).  
 GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFI135 OR TAFI130.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBITaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97336072; Pubmed=9192867;  
 RA Mengus G., May M., Carre L., Chambon P., Davidson I.;  
 RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s  
 of the retinoic acid, vitamin D3, and thyroid hormone receptors in  
 RT mammalian cells.";  
 RL Genes Dev. 11:1381-1395(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; Pubmed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babage A.K., Baguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Leivasalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McComachie L.J., McIay K., McMurtry A.A.,  
 RA Milne S.A., Misty D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.T., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [3]  
 RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=97098442; Pubmed=8942982;  
 RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;  
 RT "Molecular cloning and analysis of two subunits of the human TFIIID  
 RT complex: hTAIFI130 and hTAIFI100.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).  
 CC -I- FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT  
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS  
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION  
 CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.  
 CC -I- SUBUNIT: TFIIID IS COMPOSED OF TAF4 BINDING PROTEIN (TBP) AND A  
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).  
 CC -I- SUBCELLULAR LOCATION: Nuclear.  
 CC -I- SIMILARITY: BELONGS TO THE TAF2C FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: Y11354; CAA72189.1; -;  
 DR EMBL: AL137077; CAC36006.1; -;  
 DR EMBL: AL109911; CAC23212.2; -;  
 DR EMBL: U75308; AAC50901.1; -;  
 DR TRANSFAC: T02328; -;  
 DR GeneW: HGNC:11537; TAF4.  
 DR MIM: 601796; -;  
 DR InterPro: IPR003894; TAF\_hom.  
 DR SMART: SMO0549; TAFH.1.  
 KW Transcription regulation; Nuclear protein.  
 FT DOMAIN 39 42  
 FT POLY-HIS.  
 FT DOMAIN 52 57  
 FT POLY-ALA.  
 FT DOMAIN 98 101  
 FT POLY-GLY.  
 FT DOMAIN 142 148  
 FT POLY-ALA.  
 FT DOMAIN 268 275  
 FT POLY-PRO.  
 FT DOMAIN 331 337  
 FT POLY-ALA.  
 FT DOMAIN 680 683  
 FT POLY-PRO.  
 FT DOMAIN 808 813  
 FT POLY-ALA.  
 FT DOMAIN 828 831  
 FT POLY-ASP.  
 FT DOMAIN 105 117  
 FT PGPPSPRRPLVPA -> GRGLQQRGGRES  
 FT (IN REF. 3).  
 FT CONFLICT 136 136 A -> S (IN REF. 2).  
 FT CONFLICT 185 185 G -> GCG (IN REF. 2).  
 FT CONFLICT 233 264 MISSING (IN REF. 3).  
 FT CONFLICT 293 293 P -> L (IN REF. 3).  
 SO SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 GRC64;  
 Query Match 2.5%; Score 133; DB 1; Length 1083;  
 Best Local Similarity 20.8%; Pred. No. 0.35; Indels 202; Gaps 28;  
 Matches 135; Conservative -----LTASLPPRRDSTEARPKRSP-----EQPIGGIRIV 178

```

Db 400 KGAAGAVT-QSLSRTPATTSGIRATLTPVLAPRLPQPQNTNIONFLPGOMLVRS 458
Qy 179 GTQLRLGPEDDLAGMFLQIF-----PLSPDPMOSSPPRALALQALQGE-LARVQ 232
Db 459 ENGQLMIPQALAOAOAHAPOTTMARPTPTSAF-PVOISTYQAGCTPIARQVT 517
Qy 233 GSEPEPGIT--VVALQALATLSSPH-----GALVMSHSHFLACPLRLQOLCOYOR 283
Db 518 PTTIIRQVSOAQCTVOPSAFLQRSFGQPOLVVGAA--OTASIGTATAVOTGTPOR 572
Qy 284 CVPQDNGFSSL-----FLKVLLOMLQWIDPVEGGLPRAOLRLMASQASAG 350
Db 573 TPGATTTSSATETMENKCKKNFLSTLIK-----LAASSKOSTETRAANKELVQNLIDG 628
Qy 331 R-RLSDVRCGLLR-----LAEALAFRODLE----- 354
Db 629 KIEAEFTSLRYELNASSPQRYLVPLKRLPALRQULTPSAFLIQSOQOPPPPTSQAT 688
Qy 355 -----VVSSTVA--AVIATLRSGEQSV--EPDLISKVLQG----- 386
Db 689 TALTAVLSSSVORTAGKTAAVTASALQPPVLSLTQPTQVGVKCGOPTEPLVIQPPKPG 748
Qy 387 -----LIEVSPHLELTAFESATADAASFPACKPVVYVSSLLQEEEP 432
Db 749 ALIRPQVTLTQTPVVALKOPHNRIMLTTPQOIQLNPLQVPPVAKPVALGCTALSAVSA 808
Qy 433 LAGG-----KPGA-----DGSLEAVRLGSPSGILVWLEMLDPEVYSSCPD- 474
Db 809 QAAAGAKNKLKEPFGGSGFRDDDDINDVASMAGVNLSEASRIATNSELVGTILRSCKDE 868
Qy 475 -----LQRL-L-FSRKRGKGAQVPSFPPYLLTLTTHOSSMPTLHOCIRVLGK---- 522
Db 869 TELLPALQRLRIETLICKKHG-----ITELHPDVSVYSNAT-----QORLONIVEKISET 918
Qy 523 -----SREORFSPASLDFLMACIHYPRIMOGDOPRPOKREELV-----L 564
Db 919 AOKNFSTKDDDRYEQASVDRAOLKEF-----EQDDIQKQKKDEDERITILRAKRSRS 972
Qy 565 RVQGPPLISL---VELILAE-AETRSQGDGTACSLIQARLPLLLSC 607
Db 973 RQDPEDQLRIKQAKEMOQOELAQMRORDANTALAIIGRKKRKRYDC 1020

```

RESULT 6

PCN2\_HUMAN STANDARD; PRT; 3321 AA.

AC 095613; 043152;

DT 30-MAY-2000 (Rel. 39, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Pericentrin 2 (Kendrin).

GN PCN2 OR KIAA0402.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Li O., Joshi H.C.;

RT "Kendrin, an integral component of centrosome";

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1512-3321 FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=98116655; PubMed=9455477;

RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. VIII. 78 new cDNA clones from brain which code for large proteins in vitro.";

RL DNA Res. 4:307-313(1997).

RN [3]

```

RP REVISIONS, AND SEQUENCE OF 1192-3321 FROM N.A.
RA Ohara O.
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 787-1522 FROM N.A.
RC TISSUE=Trachea;
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N., Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y., Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S., Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R., Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y., Matanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Iisaga T., "MEDO human cDNA sequencing project.";
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: CENTROSOMAL PROTEIN.
CC -1- SIMILARITY: STRONG, TO MOUSE PERICENTRIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).


DR EMBL: U52962; AAD10838.1; -



DR EMBL: AB007862; BAA23698.2; ALT_INIT.



DR EMBL: AK093923; BAC04252.1; ALT_INIT.



DR Genew: HGNC:16068; PCNT2.



DR MIM: 605925; -



KW coiled coil.



FT DOMAIN 258 553 COILED COIL (POTENTIAL).



FT 675 835 COILED COIL (POTENTIAL).



FT 999 1135 COILED COIL (POTENTIAL).



FT DOMAIN 1288 1938 COILED COIL (POTENTIAL).



FT 2053 2071 COILED COIL (POTENTIAL).



FT DOMAIN 2525 3075 COILED COIL (POTENTIAL).



FT 819 819 A -> G (IN REF. 4).



FT 879 879 A -> T (IN REF. 4).



FT 891 891 Q -> Q (IN REF. 4).



FT 956 956 T -> S (IN REF. 4).



FT 1013 1013 K -> E (IN REF. 4).



FT 1025 1025 T -> A (IN REF. 4).



FT 1276 1276 I -> L (IN REF. 1).



FT 1306 1306 K -> T (IN REF. 1).



FT 1523 1523 H -> Q (IN REF. 2).



FT 2177 2177 R -> M (IN REF. 2).



FT 2538 2538 T -> A (IN REF. 2).



FT 2906 2906 MISSING (IN REF. 2).



FT 3125 3125 S -> Y (IN REF. 2).



FT 3289 3289 G -> E (IN REF. 2).



FT 3309 3321 YVQILQNRNPATR -> VLPDSYSKSKSCHPMIKQ (IN REF. 2).



SO SEQUENCE 3321 AA; 376355 MW; AOB693FACADAL151 CRC64;



Query Match 2.4%; Score 127; DB 1; Length 3321;



Best Local Similarity 20.4%; Pred. No. 4.5; 330; Indels 284; Gaps 39;



Matches 187; Conservative 117; Mismatches 117;



Qy 17 PPRADSEFOALDIWPEEKPLPTAFVDTSEALLPDMKLRLMRISVLRVDAALQ 76



Db 2411 PPRKED-ELID-LSINGKQGEVPTA-----CPWM-----RGDLLOVQGEAF 2451



Qy 77 -DLEPQQLLFVOSFGIVSSMSKLLQFLDQAVAHDPQTLQENIMDKNIYAH- 128



Db 2452 KEQEMOGVELQPRLSGSDLGSHSLERLEKIR-----EQGDLQEKSLHRLPDRSS 2505



Qy 129 -----VAVQHERGASGGOTFHSLLTALPPRDSSTEARPKRSSPEQPIGGRR 177


```

```

Db 2506 LLESIQALRAOLRMTHLNOEKLQHLRTALT-----STBA-----RGSQOQH-----Q 2548
Oy 178 VGTOLRVGLPEDDLAGMFLQIFPLSPDRWSSSPRPVALAQAOLG--QELARVYQSGP 235
Db 2549 LRRQVELLA-----YKVEOEKCIAGDIQKTLSEOEKANSVQKLL 2588
Oy 236 EVPGITVAVLQALATLSSPHGALVMSNHRSHFLACPLRLQDCQYORCPVDOTGSSIF 295
Db 2589 AAEQVNVVDLKS-----DLCESSQKSEQLS-----RSLCEVQGVLDLRSLSSK 2633
Oy 296 LKVLQMLQWLDSPVEGGPLRA-----QLRMLASQASGRRLSDVRGL-----LRL 343
Db 2634 ENELAAALOELESQEGKQALQSOLEEEOQLRHLORESQAKALEELRASLETQRAOSSRL 2693
Oy 344 AEAALFRQ-----DLEVSSYTRAVIATIRSGEQCEVEDLISKVQGLIEVRSPL 395
Db 2694 CVALKHEDQAKNDLOKELEIHSRCEALIAQERS--QLSELQDLAAEKSRTELESEALRH 2752
Oy 396 BELLAFFSATADASPFPACKPRVVVSSLLQEEEPILAGKRGADGQSLAVRLCPSSG 455
Db 2753 EKLTLQESQRTQEQACVHDQ-----AHNALLQ-----KLKEKS 2788
Oy 456 LLYVWLEMLDPEVSSCPDLQRLLEFSRRKGGQAQVPSFRPYLLTLFTHOSSWPTLLHC 515
Db 2789 RYVDLOAME-----KVQOQALHSQOQLEBAEQ-----KHC 2819
Oy 516 IFVLLKSGSEQRFPDSASIDFLMACIHVRIMQGRDQRTPOKRRRELVLRVQGPILISLV 575
Db 2820 -----EALRRKEVSATLKTSTVEALHT--OKRELRSLEERERKPAWLD-----A 2862
Oy 576 ELILAEAETRSQDGTAAACSLIQARPLLLSCCGDSEVRRVTEHLSGICIQONGSVLG 635
Db 2863 ELEQSHPRKKEQEGKKAARSAEAR-----QSPAADQMKR--W 2899
Oy 636 RRCRDLILQLYLQRP-----ELRVVPVEVLLHSEGAASSVCKLDGLIH 679
Db 2900 QKDKERLELELQORDLHKIKOLOQTVVRDLSEKDEVPGRSLHLGARSARAASDADHL-- 2957
Oy 680 RRTTLADTSRLEENRGADASMACRKLAVAHPLLLRHLMIALALLHGRHNLNPOERR 739
Db 2958 -----REOQRELE-----AMRORLSAA-----RLTSFTSQAVDRT-VNDWISS 2996
Oy 740 QONHLSCFLVHGLL--ELLQPHVRSRHOHALMDCLSFIRLLNYRKSRRHLAFLNK 797
Db 2997 NEKAVMSLHTLEELKSDLSRP--TSQKMAAELOFPFVYLL--KDNVSLTKALST 3050
Oy 798 FVQFIHKYITVAPPAISFLQKHADPLHDLSPDNSDVLMLKSLLAGLSLPSDDRTDGL 857
Db 3051 VTQ--EKLELSRAVSKLEKILKH-----HLOKGS--PSRSERSAMPK 3089
Oy 858 DEEGEES-----SAGSLP 871
Db 3090 DETAPOSSLRRDPGRPL 3107

```

## RESULT 7

SBCC\_PSEAE

ID SBCC\_PSEAE STANDARD: PRT: 1211 AA.

```

AC Q9HMB8.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclease sbccd subunit C.
OS SBCC OR PA4282.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RX [1]
RP SEQUENCE FROM N. A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;

```

```

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Iatibig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RA opportunistic pathogen."
RA Nature 406:959-964(2000).
CC - FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (By similarity).
CC - SUBUNIT: Heterodimer of sbcc and sbcd (By similarity).
CC - SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC DR EMBL: AE004844; AAC07670.1;
CC Interpro: IPR003439; ABC_transport.
CC
CC Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;
CC NP_BIND 37 44 ATP (POTENTIAL).
CC FT DOMAIN 115 156 COILED COIL (POTENTIAL).
CC FT DOMAIN 197 407 COILED COIL (POTENTIAL).
CC FT DOMAIN 497 584 COILED COIL (POTENTIAL).
CC FT DOMAIN 628 1048 COILED COIL (POTENTIAL).
CC FT SEQUENCE 1211 AA; 138959 MW; FC8DCA656C08E42 CRC64;

```

```

Query Match 2.4%; Score 125.5; DB 1; Length 1211;
Best Local Similarity 21.0%; Pred. No. 1.3;
Matches 232; Conservative 118; Mismatches 426; Indels 329; Gaps 49;

```

```

Oy 24 EFOALDI-----WPEERPLPAFLVDTSEELPLPMKLMISEVLRVDAALQDL 78
Db 251 EFOALQRLQEGQOOWFTEQOR-----LLQSCENH-----QGLAEARQAMDALATER 296
Oy 79 EPOQLLFQVSGFIPVSSSKLQFLQDAVANDPOTLEONIMDKNMAHLYEVOHERGAS 138
Db 297 ETLQWLERL-----APYKGLIERLKQLEQLRHSQOQR-----RTQQA 338
Oy 139 GGQTFHSLTASLPPRRDSTEAPKPSPEOPRIGGRINVTOL-----RVLGPEDD 190
Db 339 GTERLQGLQARLQEARERQADNHLQADAP-----LREARQLSEARRLERTTLAEPOE 393
Oy 191 LAGMFLQIFPLSPDRWSSSPRPVALAQAOLQGLARVAVQSGSPVPGITVAVLQALAT 250
Db 394 L-----HROSQNRHAQSDAARQIDMEQQRHVAEQDLQ--ALRDSQALAA 438
Oy 251 LLS--PHGALVMSNHRSHFLACPLRLQDCQYORCPVDOTGSSIFLVL-----LOML 303
Db 439 LGDAWYTHOGQALATYQRR-----GRALQSOAQLPELEKSLAHNGPELERL 484
Oy 304 --QWLDSPVEGGPLRAQLRLMLASQASAGR-----LSVNGGLRLAEALAFRO 351
Db 485 QAWMTALHSEPDLLAARLVELLRQDLSERQDALHKEWQVLDQAGALARRGEL----- 540
Oy 352 DLEVSSYTRAVIATIRSGEQCEVEDLISKVQGLIEV-----RSHLEELLAFFS 404
Db 541 DGRWEOE--QALLDLKRGSSQCAEEVKAADQALQVTRILLQRLRLARSASVEQLRAGLYD 599
Oy 405 ATADAASPPACKPRVVVSSLLQEEEPILAGKRGADGQSLAVRLCPSSGLVLDLEML 464
Db 600 GER-----CPVCG-----SQENP-----YHSEQLLAALGEHD 627
Oy 465 DPEVSSCPDLQ-----LRLIFS--RRKGGQAQVPSFRPYLLTLFTHOSSWPTLLH 513

```

```

DB 628 DOQVRAEDSLERLQTLVGLREGYSQREBLNQSROEOELTQGLALROLDDW-TLP 686
OY 514 QCEIRVLLGKREORFDPDSASLDFIMACIHPVPRIMQGRDPTP-OKRAEVLVRQSGELI 572
DB 687 EELRL-----QPSADLEWL-----AORLDLLAGQKOCORPDRILAROTQOL 732
OY 573 SLVELLAEETRSQDDTACSLIQAIRPLLLSCCGDESVARKV--TEHLSGCIQOW- 629
DB 733 ---QOEIRAEETLQORQALTEBQV-RYEHLLQOOVEEDSQOLRPLLSDEHW-----QRMQ 784
OY 630 -----GDSVLGRCRDLLQTLQVLR-PELRVPE-----VLHSEG 665
DB 785 ADPLRTFOALGESIEORRQOARLQOIEORLEKORCDSSWQKOSDORNEARQAE 844
OY 666 AASSVCKLGLDIHRTFTLLADTSD-SRALENRCADSMACRKLAVAHPLLLHPLMIA 724
DB 845 RAQAEELAEELGRGLAHGACACADQWOLSLFHAAQAQSAVETIQV--PLDSLREBOL-- 900
OY 725 ALLHGRTHLNFQEFROQNL-----SCFLHVLGL-----L 754
DB 901 RLAEALHELQOORROQDFORLQADQWREBODNLDSRLDALGLSEBOATQWREOL 960
OY 755 ELLOPHFR-----SEHGLMDCLSFIR-----LLNKRKSRHLAFLINFKVQ 800
DB 961 QRIQEEITRQOQLEAEKQAO---LQHRROREPTEDEALQDNRQERERLAASEQAYLE 1016
OY 801 FHKYITYNAPAAISFLQKADPLHLDSLFPNSDLYMLKSLAGLSLPSRDRTRGDEE 860
DB 1017 -----TYS-----QIADNQRREQOALLAEERARERFRGRGLNE- 1053
OY 861 GEEBSAGSLPLVSVSLFTPLTAEMAPYMKRUSRGOTVEDLEVLSDID--EKSRR-- 915
DB 1054 -----LIGSSSGDKFRRIAGQYNL-DLLVQHSNVOLRQIARRYRL 1092
OY 916 -----RPELISFSTNLQRLMSAECCCNLAFLSRMSQNSP-SIAA 958
DB 1093 QRGSEGLLVDTKEMDELSRYS-----LSGGEFLISLALGLASWASKLRIES 1146
OY 959 AFLPTFMYCLGSDQFEVYOTALRNL 983
DB 1147 LFIDEGFGLDPESLQAMDALDNL 1171

```

RESULT 8  
FACAL\_HUMAN STANDARD: PRT; 1455 AA.  
ID FACAL\_HUMAN 015360; 092497; 075266; 09Y6M2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 15-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Fanconi anemia group A protein (FACA protein).  
GN FANCA OR FAA.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97312685; PubMed=9169126;  
RA Ianzano L., D'Apolito M., Centra M., Savino M., Levay O.,  
RA Auerbach A.D., Gleason-Jensen A.M., Doggett N.A., Pronk J.C.,  
RA Callen F.C., Zelante L., Savoia A.,  
RT "The genomic organization of the Fanconi anemia group A (FAA) gene";  
RL Genomics 41:309-314(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymphoblast;  
RX MEDLINE=97051928; PubMed=8896563;  
RA Lo Ren Foe J.R., Roolmans M.A., Bosnoyan-Collins L., Alon N.,  
RA Wijket M., Parker L., Lightfoot J., Carreau M., Callen D.F.,  
RA Savoia A., Cheng N.C., van Berkel C.G.M., Strunk M.H.P., Gilie J.J.P.,

```

RA Pals G., Kruyt F.A.E., Pronk J.C., Arwert F., Buchwald M., Joenje H.;  
RT "Expression cloning of a cDNA for the major Fanconi anemia gene,  
RT FAA";  
RL Nat. Genet. 14:320-323(1996).  
RN [3]  
RP SEQUENCE OF 175-834 FROM N.A.  
RA Rieke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,  
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,  
RA Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S.,  
RA Ung S., Tatum O., Campbell C., Fawcett J., Maltbie M., Deaven L.;  
RT "Sequencing of human Fanconi anemia complementation group A gene  
RT genomic region";  
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 491-542 FROM N.A.  
RX MEDLINE=98375690; PubMed=9711872;  
RA Levran O., Doggett N.A., Auerbach A.D.;  
RT "Identification of Alu-mediated deletions in the Fanconi anemia gene  
RT FAA";  
RL Hum. Mutat. 12:145-152(1998).  
RN [5]  
RP SUBCELLULAR LOCATION, AND MUTAGENESIS.  
RX MEDLINE=98414611; PubMed=9742112;  
RA Naef D., Kupfer G.M., Suliman A., Lambert K., D'Andrea A.D.;  
RT "Functional activity of the Fanconi anemia protein FAA requires FAC  
RT binding and nuclear localization";  
RL Mol. Cell. Biol. 18:5952-5960(1998).  
RN [6]  
RP PHOSPHORYLATION.  
RX MEDLINE=99007271; PubMed=9789045;  
RA Yamashita T., Kupfer G.M., Naef D., Suliman A., Joenje H., Asano S.,  
RA D'Andrea A.D.;  
RT "The Fanconi anemia pathway requires FAA phosphorylation and FAA/FAC  
RT nuclear accumulation";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:13085-13090(1998).  
RN [7]  
RP VARIANTS FA, AND VARIANTS.  
RX MEDLINE=98058767; PubMed=9371798;  
RA Levran O., Erlich T., Magdalena N., Gregory J.J., Batish S.D.,  
RA Verlander P.C., Auerbach A.D.;  
RT "Sequence variation in the Fanconi anemia gene FAA";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:13051-13056(1997).  
RN [8]  
RP VARIANTS FA PRO-1110 AND GLY-1117.  
RX MEDLINE=99225041; PubMed=10210316;  
RA Kupfer G., Naef D., Garcia-Higuera I., Masik J., Cheng A.,  
RA Yamashita T., Tipping A., Morgan N., Mathew C.G., D'Andrea A.D.;  
RT "A patient-derived mutant form of the Fanconi anemia protein, FANCA,  
RT is defective in nuclear accumulation";  
RL Exp. Hematol. 27:587-593(1999).  
CC -1- FUNCTION: DNA REPAIR PROTEIN THAT MAY OPERATE IN A POSTREPLICATION  
CC REPAIR OR A CELL CYCLE CHECKPOINT FUNCTION. MAY BE IMPLICATED IN  
CC INTERSTRAND DNA CROSS-LINK REPAIR AND IN THE MAINTENANCE OF NORMAL  
CC CHROMOSOME STABILITY.  
CC -1- SUBUNIT: BELONGS TO A MULTISUBUNIT COMPLEX COMPOSED OF FANCA,  
CC FANCB, AND FANCG PROTEINS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (MAJOR) AND CYTOPLASMIC (MINOR).  
CC -1- PTM: PHOSPHORYLATED PRIMARILY ON SERINE RESIDUES. PHOSPHORYLATION  
CC IS REQUIRED FOR THE FORMATION OF THE NUCLEAR COMPLEX. NOT  
CC PHOSPHORYLATED IN CELLS DERIVED FROM GROUPS A, B, C, E, F, G, AND  
CC H.  
CC -1- DISEASE: DEFECTS IN FANCA ARE THE CAUSE OF ONE OF THE 8  
CC COMPLEMENTATION GROUPS OF FANCONI ANEMIA (FA), AN AUTOSOMAL  
CC RECESSIVE DISORDER CHARACTERIZED BY PROGRESSIVE PANCYTOPENIA, A  
CC DIVERSE ASSORTMENT OF CONGENITAL MALFORMATIONS, AND A  
CC PREDISPOSITION TO THE DEVELOPMENT OF MALIGNANCIES. AT THE CELLULAR  
CC LEVEL IT IS ASSOCIATED WITH HYPERSENSITIVITY TO DNA-DAMAGING  
CC AGENTS, CHROMOSOMAL INSTABILITY (INCREASED CHROMOSOME BREAKAGE),  
CC AND DEFECTIVE DNA REPAIR. FANCA DEFECTS ACCOUNT FOR 65 % OF FA.  
CC DATABASE: NAME=Fanconi Anemia Mutation database;  
CC WWW="http://www.rockefeller.edu/fanconi/mutate/";  
CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Hematol.;  
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/FA1D102.html".

```

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: 283067; CAB05445.1; JOINED.  
DR EMBL: 283068; CAB05445.1; JOINED.  
DR EMBL: 283069; CAB05445.1; JOINED.  
DR EMBL: 283070; CAB05445.1; JOINED.  
DR EMBL: 283071; CAB05445.1; JOINED.  
DR EMBL: 283072; CAB05445.1; JOINED.  
DR EMBL: 283073; CAB05445.1; JOINED.  
DR EMBL: 283074; CAB05445.1; JOINED.  
DR EMBL: 283075; CAB05445.1; JOINED.  
DR EMBL: 283076; CAB05445.1; JOINED.  
DR EMBL: 283077; CAB05445.1; JOINED.  
DR EMBL: 283078; CAB05445.1; JOINED.  
DR EMBL: 283079; CAB05445.1; JOINED.  
DR EMBL: 283080; CAB05445.1; JOINED.  
DR EMBL: 283081; CAB05445.1; JOINED.  
DR EMBL: 283082; CAB05445.1; JOINED.  
DR EMBL: 283083; CAB05445.1; JOINED.  
DR EMBL: 283084; CAB05445.1; JOINED.  
DR EMBL: 283085; CAB05445.1; JOINED.  
DR EMBL: 283086; CAB05445.1; JOINED.  
DR EMBL: 283087; CAB05445.1; JOINED.  
DR EMBL: 283088; CAB05445.1; JOINED.  
DR EMBL: 283089; CAB05445.1; JOINED.  
DR EMBL: 283090; CAB05445.1; JOINED.  
DR EMBL: 283091; CAB05445.1; JOINED.  
DR EMBL: 283092; CAB05445.1; JOINED.  
DR EMBL: 283093; CAB05445.1; JOINED.  
DR EMBL: 283094; CAB05445.1; JOINED.  
DR EMBL: 283095; CAB05445.1; JOINED.  
DR EMBL: 283151; CAB05445.1; JOINED.  
DR EMBL: X99226; AAC67610.1; -.  
DR EMBL: AC005567; AAC33401.1; -.  
DR EMBL: AF054569; AAC28331.1; -.  
DR GeneW; HGNC:3582; FANCA.  
DR MIM: 227650; -.  
DR InterPro; IPR003516; Fanconia.  
DR Pfam; PF03511; Fanconia; 1.  
DR PRINTS; PR00826; FANCONIAGEN.  
KW DNA repair; Nuclear protein; Phosphorylation; Disease mutation;  
KW Polymorphism.  
FT DOMAIN 18 34 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT VARIANT 6 6 V -> D.  
FT VARIANT 8 8 /FTid-VAR\_009637.  
FT VARIANT 8 8 N -> K (IN FA A); COULD BE A  
FT POLYMORPHISM).  
FT VARIANT 181 181 A -> V (IN FA A).  
FT VARIANT 244 244 /FTid-VAR\_009639.  
FT VARIANT 252 252 L -> F (IN FA A).  
FT VARIANT 435 435 /FTid-VAR\_009640.  
FT VARIANT 492 492 D -> G (IN FA A).  
FT VARIANT 501 501 R -> C (IN FA A).  
FT VARIANT 492 492 /FTid-VAR\_009641.  
FT VARIANT 501 501 H -> R (IN FA A).  
FT VARIANT 739 739 /FTid-VAR\_009643.  
FT VARIANT 501 501 S -> G.  
FT VARIANT 739 739 /FTid-VAR\_009644.  
FT VARIANT 809 809 P -> L.  
FT VARIANT 809 809 /FTid-VAR\_009645.  
FT VARIANT 817 817 G -> D.  
FT VARIANT 817 817 /FTid-VAR\_009646.  
FT VARIANT 817 817 L -> P (IN FA A).  
FT /FTid-VAR\_009647.

FT VARIANT 845 845 L -> P (IN FA A).  
FT /FTid-VAR\_009648.  
FT VARIANT 1055 1055 R -> L (IN FA A).  
FT /FTid-VAR\_009649.  
FT VARIANT 1110 1110 H -> P (IN FA A); LOSS OF FUNCTION).  
FT /FTid-VAR\_009650.  
FT VARIANT 1117 1117 R -> G (IN FA A); LOSS OF FUNCTION).  
FT /FTid-VAR\_009651.  
FT VARIANT 1128 1128 O -> E (IN FA A).  
FT /FTid-VAR\_009652.  
FT VARIANT 1131 1131 T -> A (IN FA A).  
FT /FTid-VAR\_009653.  
FT VARIANT 1263 1263 MISSING (IN FA A).  
FT /FTid-VAR\_009654.  
FT VARIANT 1287 1287 V -> I.  
FT /FTid-VAR\_009655.  
FT VARIANT 1302 1302 W -> R (IN FA A).  
FT /FTid-VAR\_009656.  
FT VARIANT 1328 1328 T -> A.  
FT /FTid-VAR\_009657.  
FT VARIANT 1417 1417 H -> D (IN FA A).  
FT /FTid-VAR\_009658.  
FT CONFLICT 717 717 M -> I (IN REF. 2).  
SQ SEQUENCE 1455 AA; 162804 MW; DB43EBDB29DF5C0E CRC64;  
Query Match 2.4%; Score 123.5; DB 1; Length 1455;  
Best Local Similarity 21.2%; Pred. No. 2.4; Indels 369; Gaps 63;  
Matches 256; Conservative 133; Mismatches 447;  
7 HAWVILLTLGPPRRADSEFOALLDIFPEPKPLPTA-----FLVDTSEAL-LLPDLKL 60  
322 HTLTQITSPVLKASDAVOMQREMSFARHPLTLTLRYLFVMAAEIVGHLOEVLE 381  
61 RMIR-SEVLRVDAQLD-EPQQL-----LPVQSF-GIPVSSKLLQFLDQAVAND 111  
382 QEVHWQVLSFEVALVCEPEAOQLLEDVVARLMAOAFESCQDSVTAFLVVRQALEG 441  
112 PQLTEONIMDKNMAHLEVOHERGASGO-----TFHSLTLASLPDRDSTE 159  
442 PSA-----PLSTADMFKAFCSTRGHGSKALVLFTELSLIVFESPRRLQVH 492  
160 APRKSSPEQ-----PLGGGRIVG--TQLRVGLPEDDLGAMFLQIFPLSPDRMOS 209  
493 ILHPVLVPSKYSRLTDYLSLAKTRLADLKVSLENNGLYEDL----- 534  
210 SSPRPVALALQALGOELARVYOGSEPEVGITRVLOALATLISPHGALVMSHRSH 269  
535 SSAGDITPEHSQAL-ODVERAIVFEHTGNIPVTME--ASTIRRPY-----IVSH 583  
270 LACPLRLQCYORCVPODTGFSSFLKVL-----LOMLQWLDSPGVEGGLRAQI 320  
584 LPAALL-----PRVLPKVPDSNVAFIESLKRADKLPISLYSYVCGACSAAEKPPDAL 637  
321 RMLASQAGRLSDVRGGLRLAELARODLEVVSSTYRAVIATLR-----SGQCSY 375  
638 GVAEENSAAEELPGLTALGELRASMTDPSQDRVISAQVAVISELRAVLGHNEDDSV 697  
376 EPLILSKVLQGLIEVSPHLE-----LLTAFSATADASPP----- 414  
698 E--ISKI--QUSTNTPRIEPRHMAVDLLTSCONLMAASSVAPPEBQWALLAVR 751  
415 -ACKPVV--VVSGL--LLOEEPLAGKPGADGSLAVARLGPSSGLVDWLEMLDEPV 469  
752 TMGGRVLPVLTQLCOLLRHGQSLSA-PRVLGALALAVHLGSRAL-----PEVD 802  
470 SSCPCDQLRLLESRRKGKQAOVPSFRPVYLLFTHQSSMPTLHQG--IRVLGKSREO 526  
803 VGPP-----APGAG-LPVLPALFSLTLCRTDLSLFFCLKFTAAISYSLCKFSSO 851  
527 RPDPSAS-----LDPL-----MACIHP-----RTW 547  
852 SRDTLCSCSPGLIKKFOPLMFRPLFSEAROPISSEEDVASLSMRPLTHPSADKORALSL 911



QY 548 OGRDPTPKRR-----ELVAVQPELISLV-----ELLIAEETR 585  
 Db 912 THFRFEVLEEDVHTLYQDMLHLEIQ--PEADALSDTERQDHPQAHMHFPRESS 970  
 QY 586 SQ-DGDTAAACSLIOARPLLLSCCGGDESVRYKTEHLSGCIQMGDSVLC----- 635  
 Db 971 GGCGGD-----LOACTIIVNALMDHOSRSR-DH-----SENSDLVGGRGNGEDIT 1018  
 QY 636 RRCRDLLOLYLQRPRLVPEVYLHSEGAASSVCKLDGLHRTLLADTSDSRALE 695  
 Db 1019 SRLQEMVADELQO-DLIVPL-----GHTPSQEHFLPEIFRRRLQALT-----S 1061  
 QY 696 NRGDASMACRKLAVAPRLRLHPLMIALHRTHLNQEFQOHNLSG-----FLHVL 751  
 Db 1062 GMSVAASLQRPRLMYKRLLR-LP--SSVLCG-----SSFAEQPITARCQEPFHLV 1112  
 QY 752 GLLELLQPHVFRSE-----HOGAL-WDCLLSFRLLNRYKSSRHAAAFINKVFQFIHK 804  
 Db 1113 -----NSBMRNFCSHGALITODITAHFRGLNACLNSRDPSL-----WDFILA 1157  
 QY 805 YITYNAPAAISFLQKADPHLHDSFDSNDLVMLKSLAGL-----SLPSRDRTD 854  
 Db 1158 KCQTKCPULIT-----SALVMPSLPEVILCRMRHRCQSPRLPRLQKL- 1200  
 QY 855 RGDEBESEESSAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQ--TWEDILE--VLSDI 909  
 Db 1201 -----QEGROFASD-----FLSPEASPAENPDMLSAALHFAIOOVREENIKROL 1246  
 QY 910 DEMSRRPPELISF-FSTNLORLMSAECCRNLAFLSLAKSOMSPS---IAAFLPTF- 964  
 Db 1247 KKLDCERELLVLFPPFSKLGLSS-----HLTNSITDLPFAHFCAAILLECLE 1296  
 QY 965 -----MYCLOSODEVYQVOTALRNLP-----YALLCOEHA-AVLLHRAFL- 1003  
 Db 1297 KRKISMLALPOLMESDLRLGLRLLRVAPDQTRLLPPAFVSLSYFHEDAIAIFREAFLHV 1356  
 QY 1004 -VGMV 1007  
 Db 1357 AVDMY 1361

RESULT 9  
 SPOC\_HUMAN  
 ID SPOC\_HUMAN STANDARD; PRT; 2564 AA.  
 AC Q9H254; Q9HCD0; Q9H3G8; Q9H1K7; Q9H1K8; Q9H1I9;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3)  
 DE (Beta-IV spectrin).  
 GN SPTBN4 OR SPTBN3 OR KIAA1642.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_Taxid:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=21316449; PubMed=11294830;  
 RA Tse W.T., Tang J., Jin O., Korsgren C., John K.M., Kung A.L.,  
 RA Gwyn B., Peters L.L., Lux S.E.;  
 RT "A new spectrin, beta-IV, has a major truncated isoform that  
 RT associates with promyelocytic leukemia protein nuclear bodies and the  
 RT nuclear matrix";  
 RL J. Biol. Chem. 276:23974-23985(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 TO 4).  
 RX MEDLINE=20539976; PubMed=11086001;  
 RA Bergs S., Aguijaro D., Dirix R.Jr., Maksimova E., Stabach P.,  
 RA Hermei J.-M., Zhang J.-P., Philbrick W., Slepnev V., Ott T.,  
 RA Solimena M.;  
 RT "BetaIV spectrin, a new spectrin localized at axon initial segments  
 RT and nodes of ranvier in the central and peripheral nervous system";  
 RL J. Cell Biol. 151:985-1002(2000).

RN [3]  
 RP SEQUENCE OF 386-2382 FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain;  
 RX MEDLINE=20450683; PubMed=10997877;  
 RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes.  
 RT XVIII: The complete sequences of 100 new cDNA clones from brain which  
 RT code for large proteins in vitro";  
 RL DNA Res. 7:273-281(2000).  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1 (shown here), 2, 3 and 4; are  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Abundantly expressed in brain and pancreatic  
 CC islets.  
 CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 18 SPECTRIN REPEATS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF311855; AAG42473.1; -  
 DR EMBL: AF082075; AAG38874.1; -  
 DR EMBL: AY004226; AAF93171.1; -  
 DR EMBL: AY004226; AAF93172.1; -  
 DR EMBL: AY004227; AAF93173.1; -  
 DR EMBL: AB046862; BAB13468.1; -  
 DR HSSP: Q01082; 1BKR.  
 DR Genew: HGNC:14896; SPTBN4.  
 DR MIM: 606214; -  
 DR InterPro: IPR001589; Actbind.actnin.  
 DR InterPro: IPR001715; Calponin-like.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR002017; Spectrin.  
 DR InterPro: IPR001605; Spectrin\_PH.  
 DR Pfam: PF00169; PH; 3.  
 DR Pfam: PF00307; CH; 2.  
 DR Pfam: PF00435; spectrin; 56.  
 DR PRINTS: PR00683; SPECTRINPH.  
 DR SMART: SM00033; CH; 2.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00150; SPEC; 16.  
 DR PROSITE: PS00019; ACTININ\_1; 1.  
 DR PROSITE: PS00020; ACTININ\_2; 1.  
 DR PROSITE: PS50021; CH; 2.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 DR Cytokeleton; Membrane; Repeat; Actin-binding; Capping protein;  
 KW Alternative splicing.  
 FT DOMAIN 1 282  
 FT DOMAIN 61 165  
 FT DOMAIN 180 282  
 FT REPEAT 309 354 CH 2.  
 FT REPEAT 398 419 SPECTRIN 1.  
 FT REPEAT 429 533 SPECTRIN 2.  
 FT REPEAT 535 642 SPECTRIN 3.  
 FT REPEAT 644 771 SPECTRIN 4.  
 FT REPEAT 773 879 SPECTRIN 5.  
 FT REPEAT 881 985 SPECTRIN 6.  
 FT REPEAT 1019 1086 SPECTRIN 7.  
 FT REPEAT 1088 1197 SPECTRIN 8.  
 FT REPEAT 1199 1303 SPECTRIN 9.  
 FT REPEAT 1305 1408 SPECTRIN 10.  
 FT REPEAT 1410 1513 SPECTRIN 11.  
 FT REPEAT 1515 1619 SPECTRIN 12.  
 FT REPEAT 1621 1725 SPECTRIN 13.  
 FT REPEAT 1727 1832 SPECTRIN 14.  
 FT REPEAT 1834 1940 SPECTRIN 15.  
 FT REPEAT 1942 2046 SPECTRIN 16.  
 FT REPEAT 1942 2046 SPECTRIN 17.







Db 328 LPS-----LFNPNDAFH-----EKTYSMD-----FVRRLRQCSQA 362  
 OY 614 SVRKVTEH--LSGCIQOMGDSVLRRCRDLLOLYIQ-----REPRLVPEVILLHSG 665  
 Db 363 SVKRLRAHAYHSFNKKMN-----LPVYQIRFRELASLEALDVL--EDA 408  
 OY 666 AASSVCKLDGLIHR-FITLADTSDSRALENNGADAMACKRLAVAHPLLRLHPLMIA 724  
 Db 409 PAESPCLIAS--HRTWSILRQMSDEMFL-----PLVHR-----442  
 OY 725 ALLHGTHTLNPFQFROONHLSCTPLHVLGLLELLQPHVRSRSEHOGLMDCILSTRILLNV 784  
 Db 443 -----LMRULTL---QILARY 454  
 OY 785 RKSRRLAIFINKFVQFIKITYTNAPALISFLQKHADPLHLSFNSDLVLMKSLAGI 844  
 Db 455 -----SVFVN--ELSLRPIISNEPKREK-----KPLVYTS 482  
 OY 845 SLPSRDDRTDRGLDEGESESSAGSLPLVVSLEFPLTAEMAPYMKRLSRGQTVEDLLE 904  
 Db 483 KEPS---ITQGMTEDDGSGPSE--TKPVVSI-----SRQ---LVY 515  
 OY 905 VLSDIDEMRRRPEILSFSTNLQRLMSSAECCRLNLSLRMNSPSTIAAF---L 961  
 Db 516 VVADLDKLOEOLPELLEILIKPKLEMI-----GPKNFSISALEDSQ 557  
 OY 962 PTFMYCLGSDQDFEWW 977  
 Db 558 SFSACVPSLSKTIIO 573

## RESULT 12

EVPL\_MOUSE STANDARD: PRT: 2035 AA.

AC Q9D952; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Envoplakin (p210) (210 kDa cornified envelope precursor).  
 GN EVPL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/Sv.  
 RA MEDLINE=20347896; PubMed=10747979;  
 RA Maatta A., Ruhrberg C., Watt F.M.;  
 RT "Structure and regulation of the envoplakin gene";  
 RL J. Biol. Chem. 275:19857-19865(2000).  
 RP [2]  
 RP SEQUENCE OF 1860-2035 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Pancreas;  
 RA MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Batsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher N., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,  
 RA Hayashizaki Y.;

RT "functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 CC -!- FUNCTION: COMPONENT OF THE CORNIFIED ENVELOPE OF KERATINOCYTES.  
 CC MAY LINK THE CORNIFIED ENVELOPE TO DESMOSOMES AND INTERMEDIATE  
 CC FILAMENTS.  
 CC -!- SUBUNIT: MAY FORM A HOMODIMER OR A HETERODIMER WITH PPL.  
 CC -!- SUBCELLULAR LOCATION: COLOCALIZED WITH DSP AT DESMOSOMES AND ALONG  
 CC INTERMEDIATE FILAMENTS (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 7 PLECTRIN REPEATS.  
 CC -!- SIMILARITY: CONTAINS 1 SPECTRIN REPEAT.  
 CC -!- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation at  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: AJ309317; CAC38864.2; -  
 DR EMBL: AJ319607; CAC38864.2; JOINED.  
 DR EMBL: AJ319608; CAC38864.2; JOINED.  
 DR EMBL: AJ319609; CAC38864.2; JOINED.  
 DR EMBL: AJ319610; CAC38864.2; JOINED.  
 DR EMBL: AJ319611; CAC38864.2; JOINED.  
 DR EMBL: AJ319612; CAC38864.2; JOINED.  
 DR EMBL: AJ319613; CAC38864.2; JOINED.  
 DR EMBL: AK007353; BAB24980.1; ALT\_INIT.  
 DR MGD: MG1:107507; EVPL.  
 DR InterPro: IPR001101; Plectin\_repeat.  
 DR Pfam: PF00681; Plectin; 3.  
 DR SMART: SM00250; PLEC; 8.  
 KW Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein.  
 DR  
 FT DOMAIN 1 841  
 FT DOMAIN 2 842  
 FT DOMAIN 3 1674  
 FT DOMAIN 4 1675  
 FT DOMAIN 5 1676  
 FT DOMAIN 6 1677  
 FT DOMAIN 7 1678  
 FT DOMAIN 8 1679  
 FT DOMAIN 9 1680  
 FT DOMAIN 10 1681  
 FT DOMAIN 11 1682  
 FT DOMAIN 12 1683  
 FT DOMAIN 13 1684  
 FT DOMAIN 14 1685  
 FT DOMAIN 15 1686  
 FT DOMAIN 16 1687  
 FT DOMAIN 17 1688  
 FT DOMAIN 18 1689  
 FT DOMAIN 19 1690  
 FT DOMAIN 20 1691  
 FT DOMAIN 21 1692  
 FT DOMAIN 22 1693  
 FT DOMAIN 23 1694  
 FT DOMAIN 24 1695  
 FT DOMAIN 25 1696  
 FT DOMAIN 26 1697  
 FT DOMAIN 27 1698  
 FT DOMAIN 28 1699  
 FT DOMAIN 29 1700  
 FT DOMAIN 30 1701  
 FT DOMAIN 31 1702  
 FT DOMAIN 32 1703  
 FT DOMAIN 33 1704  
 FT DOMAIN 34 1705  
 FT DOMAIN 35 1706  
 FT DOMAIN 36 1707  
 FT DOMAIN 37 1708  
 FT DOMAIN 38 1709  
 FT DOMAIN 39 1710  
 FT DOMAIN 40 1711  
 FT DOMAIN 41 1712  
 FT DOMAIN 42 1713  
 FT DOMAIN 43 1714  
 FT DOMAIN 44 1715  
 FT DOMAIN 45 1716  
 FT DOMAIN 46 1717  
 FT DOMAIN 47 1718  
 FT DOMAIN 48 1719  
 FT DOMAIN 49 1720  
 FT DOMAIN 50 1721  
 FT DOMAIN 51 1722  
 FT DOMAIN 52 1723  
 FT DOMAIN 53 1724  
 FT DOMAIN 54 1725  
 FT DOMAIN 55 1726  
 FT DOMAIN 56 1727  
 FT DOMAIN 57 1728  
 FT DOMAIN 58 1729  
 FT DOMAIN 59 1730  
 FT DOMAIN 60 1731  
 FT DOMAIN 61 1732  
 FT DOMAIN 62 1733  
 FT DOMAIN 63 1734  
 FT DOMAIN 64 1735  
 FT DOMAIN 65 1736  
 FT DOMAIN 66 1737  
 FT DOMAIN 67 1738  
 FT DOMAIN 68 1739  
 FT DOMAIN 69 1740  
 FT DOMAIN 70 1741  
 FT DOMAIN 71 1742  
 FT DOMAIN 72 1743  
 FT DOMAIN 73 1744  
 FT DOMAIN 74 1745  
 FT DOMAIN 75 1746  
 FT DOMAIN 76 1747  
 FT DOMAIN 77 1748  
 FT DOMAIN 78 1749  
 FT DOMAIN 79 1750  
 FT DOMAIN 80 1751  
 FT DOMAIN 81 1752  
 FT DOMAIN 82 1753  
 FT DOMAIN 83 1754  
 FT DOMAIN 84 1755  
 FT DOMAIN 85 1756  
 FT DOMAIN 86 1757  
 FT DOMAIN 87 1758  
 FT DOMAIN 88 1759  
 FT DOMAIN 89 1760  
 FT DOMAIN 90 1761  
 FT DOMAIN 91 1762  
 FT DOMAIN 92 1763  
 FT DOMAIN 93 1764  
 FT DOMAIN 94 1765  
 FT DOMAIN 95 1766  
 FT DOMAIN 96 1767  
 FT DOMAIN 97 1768  
 FT DOMAIN 98 1769  
 FT DOMAIN 99 1770  
 FT DOMAIN 100 1771  
 FT DOMAIN 101 1772  
 FT DOMAIN 102 1773  
 FT DOMAIN 103 1774  
 FT DOMAIN 104 1775  
 FT DOMAIN 105 1776  
 FT DOMAIN 106 1777  
 FT DOMAIN 107 1778  
 FT DOMAIN 108 1779  
 FT DOMAIN 109 1780  
 FT DOMAIN 110 1781  
 FT DOMAIN 111 1782  
 FT DOMAIN 112 1783  
 FT DOMAIN 113 1784  
 FT DOMAIN 114 1785  
 FT DOMAIN 115 1786  
 FT DOMAIN 116 1787  
 FT DOMAIN 117 1788  
 FT DOMAIN 118 1789  
 FT DOMAIN 119 1790  
 FT DOMAIN 120 1791  
 FT DOMAIN 121 1792  
 FT DOMAIN 122 1793  
 FT DOMAIN 123 1794  
 FT DOMAIN 124 1795  
 FT DOMAIN 125 1796  
 FT DOMAIN 126 1797  
 FT DOMAIN 127 1798  
 FT DOMAIN 128 1799  
 FT DOMAIN 129 1800  
 FT DOMAIN 130 1801  
 FT DOMAIN 131 1802  
 FT DOMAIN 132 1803  
 FT DOMAIN 133 1804  
 FT DOMAIN 134 1805  
 FT DOMAIN 135 1806  
 FT DOMAIN 136 1807  
 FT DOMAIN 137 1808  
 FT DOMAIN 138 1809  
 FT DOMAIN 139 1810  
 FT DOMAIN 140 1811  
 FT DOMAIN 141 1812  
 FT DOMAIN 142 1813  
 FT DOMAIN 143 1814  
 FT DOMAIN 144 1815  
 FT DOMAIN 145 1816  
 FT DOMAIN 146 1817  
 FT DOMAIN 147 1818  
 FT DOMAIN 148 1819  
 FT DOMAIN 149 1820  
 FT DOMAIN 150 1821  
 FT DOMAIN 151 1822  
 FT DOMAIN 152 1823  
 FT DOMAIN 153 1824  
 FT DOMAIN 154 1825  
 FT DOMAIN 155 1826  
 FT DOMAIN 156 1827  
 FT DOMAIN 157 1828  
 FT DOMAIN 158 1829  
 FT DOMAIN 159 1830  
 FT DOMAIN 160 1831  
 FT DOMAIN 161 1832  
 FT DOMAIN 162 1833  
 FT DOMAIN 163 1834  
 FT DOMAIN 164 1835  
 FT DOMAIN 165 1836  
 FT DOMAIN 166 1837  
 FT DOMAIN 167 1838  
 FT DOMAIN 168 1839  
 FT DOMAIN 169 1840  
 FT DOMAIN 170 1841  
 FT DOMAIN 171 1842  
 FT DOMAIN 172 1843  
 FT DOMAIN 173 1844  
 FT DOMAIN 174 1845  
 FT DOMAIN 175 1846  
 FT DOMAIN 176 1847  
 FT DOMAIN 177 1848  
 FT DOMAIN 178 1849  
 FT DOMAIN 179 1850  
 FT DOMAIN 180 1851  
 FT DOMAIN 181 1852  
 FT DOMAIN 182 1853  
 FT DOMAIN 183 1854  
 FT DOMAIN 184 1855  
 FT DOMAIN 185 1856  
 FT DOMAIN 186 1857  
 FT DOMAIN 187 1858  
 FT DOMAIN 188 1859  
 FT DOMAIN 189 1860  
 FT DOMAIN 190 1861  
 FT DOMAIN 191 1862  
 FT DOMAIN 192 1863  
 FT DOMAIN 193 1864  
 FT DOMAIN 194 1865  
 FT DOMAIN 195 1866  
 FT DOMAIN 196 1867  
 FT DOMAIN 197 1868  
 FT DOMAIN 198 1869  
 FT DOMAIN 199 1870  
 FT DOMAIN 200 1871  
 FT DOMAIN 201 1872  
 FT DOMAIN 202 1873  
 FT DOMAIN 203 1874  
 FT DOMAIN 204 1875  
 FT DOMAIN 205 1876  
 FT DOMAIN 206 1877  
 FT DOMAIN 207 1878  
 FT DOMAIN 208 1879  
 FT DOMAIN 209 1880  
 FT DOMAIN 210 1881  
 FT DOMAIN 211 1882  
 FT DOMAIN 212 1883  
 FT DOMAIN 213 1884  
 FT DOMAIN 214 1885  
 FT DOMAIN 215 1886  
 FT DOMAIN 216 1887  
 FT DOMAIN 217 1888  
 FT DOMAIN 218 1889  
 FT DOMAIN 219 1890  
 FT DOMAIN 220 1891  
 FT DOMAIN 221 1892  
 FT DOMAIN 222 1893  
 FT DOMAIN 223 1894  
 FT DOMAIN 224 1895  
 FT DOMAIN 225 1896  
 FT DOMAIN 226 1897  
 FT DOMAIN 227 1898  
 FT DOMAIN 228 1899  
 FT DOMAIN 229 1900  
 FT DOMAIN 230 1901  
 FT DOMAIN 231 1902  
 FT DOMAIN 232 1903  
 FT DOMAIN 233 1904  
 FT DOMAIN 234 1905  
 FT DOMAIN 235 1906  
 FT DOMAIN 236 1907  
 FT DOMAIN 237 1908  
 FT DOMAIN 238 1909  
 FT DOMAIN 239 1910  
 FT DOMAIN 240 1911  
 FT DOMAIN 241 1912  
 FT DOMAIN 242 1913  
 FT DOMAIN 243 1914  
 FT DOMAIN 244 1915  
 FT DOMAIN 245 1916  
 FT DOMAIN 246 1917  
 FT DOMAIN 247 1918  
 FT DOMAIN 248 1919  
 FT DOMAIN 249 1920  
 FT DOMAIN 250 1921  
 FT DOMAIN 251 1922  
 FT DOMAIN 252 1923  
 FT DOMAIN 253 1924  
 FT DOMAIN 254 1925  
 FT DOMAIN 255 1926  
 FT DOMAIN 256 1927  
 FT DOMAIN 257 1928  
 FT DOMAIN 258 1929  
 FT DOMAIN 259 1930  
 FT DOMAIN 260 1931  
 FT DOMAIN 261 1932  
 FT DOMAIN 262 1933  
 FT DOMAIN 263 1934  
 FT DOMAIN 264 1935  
 FT DOMAIN 265 1936  
 FT DOMAIN 266 1937  
 FT DOMAIN 267 1938  
 FT DOMAIN 268 1939  
 FT DOMAIN 269 1940  
 FT DOMAIN 270 1941  
 FT DOMAIN 271 1942  
 FT DOMAIN 272 1943  
 FT DOMAIN 273 1944  
 FT DOMAIN 274 1945  
 FT DOMAIN 275 1946  
 FT DOMAIN 276 1947  
 FT DOMAIN 277 1948  
 FT DOMAIN 278 1949  
 FT DOMAIN 279 1950  
 FT DOMAIN 280 1951  
 FT DOMAIN 281 1952  
 FT DOMAIN 282 1953  
 FT DOMAIN 283 1954  
 FT DOMAIN 284 1955  
 FT DOMAIN 285 1956  
 FT DOMAIN 286 1957  
 FT DOMAIN 287 1958  
 FT DOMAIN 288 1959  
 FT DOMAIN 289 1960  
 FT DOMAIN 290 1961  
 FT DOMAIN 291 1962  
 FT DOMAIN 292 1963  
 FT DOMAIN 293 1964  
 FT DOMAIN 294 1965  
 FT DOMAIN 295 1966  
 FT DOMAIN 296 1967  
 FT DOMAIN 297 1968  
 FT DOMAIN 298 1969  
 FT DOMAIN 299 1970  
 FT DOMAIN 300 1971  
 FT DOMAIN 301 1972  
 FT DOMAIN 302 1973  
 FT DOMAIN 303 1974  
 FT DOMAIN 304 1975  
 FT DOMAIN 305 1976  
 FT DOMAIN 306 1977  
 FT DOMAIN 307 1978  
 FT DOMAIN 308 1979  
 FT DOMAIN 309 1980  
 FT DOMAIN 310 1981  
 FT DOMAIN 311 1982  
 FT DOMAIN 312 1983  
 FT DOMAIN 313 1984  
 FT DOMAIN 314 1985  
 FT DOMAIN 315 1986  
 FT DOMAIN 316 1987  
 FT DOMAIN 317 1988  
 FT DOMAIN 318 1989  
 FT DOMAIN 319 1990  
 FT DOMAIN 320 1991  
 FT DOMAIN 321 1992  
 FT DOMAIN 322 1993  
 FT DOMAIN 323 1994  
 FT DOMAIN 324 1995  
 FT DOMAIN 325 1996  
 FT DOMAIN 326 1997  
 FT DOMAIN 327 1998  
 FT DOMAIN 328 1999  
 FT DOMAIN 329 2000  
 FT DOMAIN 330 2001  
 FT DOMAIN 331 2002  
 FT DOMAIN 332 2003  
 FT DOMAIN 333 2004  
 FT DOMAIN 334 2005  
 FT DOMAIN 335 2006  
 FT DOMAIN 336 2007  
 FT DOMAIN 337 2008  
 FT DOMAIN 338 2009  
 FT DOMAIN 339 2010  
 FT DOMAIN 340 2011  
 FT DOMAIN 341 2012  
 FT DOMAIN 342 2013  
 FT DOMAIN 343 2014  
 FT DOMAIN 344 2015  
 FT DOMAIN 345 2016  
 FT DOMAIN 346 2017  
 FT DOMAIN 347 2018  
 FT DOMAIN 348 2019  
 FT DOMAIN 349 2020  
 FT DOMAIN 350 2021  
 FT DOMAIN 351 2022  
 FT DOMAIN 352 2023  
 FT DOMAIN 353 2024  
 FT DOMAIN 354 2025  
 FT DOMAIN 355 2026  
 FT DOMAIN 356 2027  
 FT DOMAIN 357 2028  
 FT DOMAIN 358 2029  
 FT DOMAIN 359 2030  
 FT DOMAIN 360 2031  
 FT DOMAIN 361 2032  
 FT DOMAIN 362 2033  
 FT DOMAIN 363 2034  
 FT DOMAIN 364 2035  
 FT DOMAIN 365 2036  
 FT DOMAIN 366 2037  
 FT DOMAIN 367 2038  
 FT DOMAIN 368 2039  
 FT DOMAIN 369 2040  
 FT DOMAIN 370 2041  
 FT DOMAIN 371 2042  
 FT DOMAIN 372 2043  
 FT DOMAIN 373 2044  
 FT DOMAIN 374 2045  
 FT DOMAIN 375 2046  
 FT DOMAIN 376 2047  
 FT DOMAIN 377 2048  
 FT DOMAIN 378 2049  
 FT DOMAIN 379 2050  
 FT DOMAIN 380 2051  
 FT DOMAIN 381 2052  
 FT DOMAIN 382 2053  
 FT DOMAIN 383 2054  
 FT DOMAIN 384 2055  
 FT DOMAIN 385 2056  
 FT DOMAIN 386 2057  
 FT DOMAIN 387 2058  
 FT DOMAIN 388 2059  
 FT DOMAIN 389 2060  
 FT DOMAIN 390 2061  
 FT DOMAIN 391 2062  
 FT DOMAIN 392 2063  
 FT DOMAIN 393 2064  
 FT DOMAIN 394 2065  
 FT DOMAIN 395 2066  
 FT DOMAIN 396 2067  
 FT DOMAIN 397 2068  
 FT DOMAIN 398 2069  
 FT DOMAIN 399 2070  
 FT DOMAIN 400 2071  
 FT DOMAIN 401 2072  
 FT DOMAIN 402 2073  
 FT DOMAIN 403 2074  
 FT DOMAIN 404 2075  
 FT DOMAIN 405 2076  
 FT DOMAIN 406 2077  
 FT DOMAIN 407 2078  
 FT DOMAIN 408 2079  
 FT DOMAIN 409 2080  
 FT DOMAIN 410 2081  
 FT DOMAIN 411 2082  
 FT DOMAIN 412 2083  
 FT DOMAIN 413 2084  
 FT DOMAIN 414 2085  
 FT DOMAIN 415 2086  
 FT DOMAIN 416 2087  
 FT DOMAIN 417 2088  
 FT DOMAIN 418 2089  
 FT DOMAIN 419 2090  
 FT DOMAIN 420 2091  
 FT DOMAIN 421 2092  
 FT DOMAIN 422 2093  
 FT DOMAIN 423 2094  
 FT DOMAIN 424 2095  
 FT DOMAIN 425 2096  
 FT DOMAIN 426 2097  
 FT DOMAIN 427 2098  
 FT DOMAIN 428 2099  
 FT DOMAIN 429 2100  
 FT DOMAIN 430 2101  
 FT DOMAIN 431 2102  
 FT DOMAIN 432 2103  
 FT DOMAIN 433 2104  
 FT DOMAIN 434 2105  
 FT DOMAIN 435 2106  
 FT DOMAIN 436 2107  
 FT DOMAIN 437 2108  
 FT DOMAIN 438 2109  
 FT DOMAIN 439 2110  
 FT DOMAIN 440 2111  
 FT DOMAIN 441 2112  
 FT DOMAIN 442 2113  
 FT DOMAIN 443 2114  
 FT DOMAIN 444 2115  
 FT DOMAIN 445 2116  
 FT DOMAIN 446 2117  
 FT DOMAIN 447 2118  
 FT DOMAIN 448 2119  
 FT DOMAIN 449 2120  
 FT DOMAIN 450 2121  
 FT DOMAIN 451 2122  
 FT DOMAIN 452 2123  
 FT DOMAIN 453 2124  
 FT DOMAIN 454 2125  
 FT DOMAIN 455 2126  
 FT DOMAIN 456 2127  
 FT DOMAIN 457 2128  
 FT DOMAIN 458 2129  
 FT DOMAIN 459 2130  
 FT DOMAIN 460 2131  
 FT DOMAIN 461 2132  
 FT DOMAIN 462 2133  
 FT DOMAIN 463 2134  
 FT DOMAIN 464 2135  
 FT DOMAIN 465 2136  
 FT DOMAIN 466 2137  
 FT DOMAIN 467 2138  
 FT DOMAIN 468 2139  
 FT DOMAIN 469 2140  
 FT DOMAIN 470 2141  
 FT DOMAIN 471 2142  
 FT DOMAIN 472 2143  
 FT DOMAIN 473 2144  
 FT DOMAIN 474 2145  
 FT DOMAIN 475 2146  
 FT DOMAIN 476 2147  
 FT DOMAIN 477 2148  
 FT DOMAIN 478 2149  
 FT DOMAIN 479 2150  
 FT DOMAIN 480 2151  
 FT DOMAIN 481 2152  
 FT DOMAIN 482 2153  
 FT DOMAIN 483 2154  
 FT DOMAIN 484 2155  
 FT DOMAIN 485 2156  
 FT DOMAIN 486 2157  
 FT DOMAIN 487 2158  
 FT DOMAIN 488 2159  
 FT DOMAIN 489 2160  
 FT DOMAIN 490 2161  
 FT DOMAIN 491 2162  
 FT DOMAIN 492 2163  
 FT DOMAIN 493 2164  
 FT DOMAIN 494 2165  
 FT DOMAIN 495 2166  
 FT DOMAIN 496 2167  
 FT DOMAIN 497 2168  
 FT DOMAIN 498 2169  
 FT DOMAIN 499 2170  
 FT DOMAIN 500 2171  
 FT DOMAIN 501 2172  
 FT DOMAIN 502 2173  
 FT DOMAIN 503 2174  
 FT DOMAIN 504 2175  
 FT DOMAIN 505 2176  
 FT DOMAIN 506 2177  
 FT DOMAIN 507 2178  
 FT DOMAIN 508 2179  
 FT DOMAIN 509 2180  
 FT DOMAIN 510 2181  
 FT DOMAIN 511 2182  
 FT DOMAIN 512 2183  
 FT DOMAIN 513 2184  
 FT DOMAIN 514 2185  
 FT DOMAIN 515 2186  
 FT DOMAIN 516 2187  
 FT DOMAIN 517 2188  
 FT DOMAIN 518 2189  
 FT DOMAIN 519 2190  
 FT DOMAIN 520 2191  
 FT DOMAIN 521 2192  
 FT DOMAIN 522 2193  
 FT DOMAIN 523 2194  
 FT DOMAIN 524 2195  
 FT DOMAIN 525 2196  
 FT DOMAIN 526 2197  
 FT DOMAIN 527 2198  
 FT DOMAIN 528 2199  
 FT DOMAIN 529 2200  
 FT DOMAIN 530 2201  
 FT DOMAIN 531 2202  
 FT DOMAIN 532 2203  
 FT DOMAIN 533 2204  
 FT DOMAIN 534 2205  
 FT DOMAIN 535 2206  
 FT DOMAIN 536 2207  
 FT DOMAIN 537 2208  
 FT DOMAIN 538 2209  
 FT DOMAIN 539 2210  
 FT DOMAIN 540 2211  
 FT DOMAIN 541 2212  
 FT DOMAIN 542 2213  
 FT DOMAIN 543 2214  
 FT DOMAIN 544 2215  
 FT DOMAIN 545 2216  
 FT DOMAIN 546 2217  
 FT DOMAIN 547 2218  
 FT DOMAIN 548 2219  
 FT DOMAIN 549 2220  
 FT DOMAIN 550 2221  
 FT DOMAIN 551 2222  
 FT DOMAIN 552 2223  
 FT DOMAIN 553 2224  
 FT DOMAIN 554 2225  
 FT DOMAIN 555 2226  
 FT DOMAIN 556 2227  
 FT DOMAIN 557 2228  
 FT DOMAIN 558 2229  
 FT DOMAIN 559 2230  
 FT DOMAIN 560 2231  
 FT DOMAIN 561 2232  
 FT DOMAIN 562 2233  
 FT DOMAIN 563 2234  
 FT DOMAIN 564 2235  
 FT DOMAIN 565 2236  
 FT DOMAIN 566 2237  
 FT DOMAIN 567 2238  
 FT DOMAIN 568 2239  
 FT DOMAIN 569 2240  
 FT DOMAIN 570 2241  
 FT DOMAIN 571 2242  
 FT DOMAIN 572 2243  
 FT DOMAIN 573 2244  
 FT DOMAIN 574 2245  
 FT DOMAIN 575 2246  
 FT DOMAIN 576 2247  
 FT DOMAIN 577 2248  
 FT DOMAIN 578 2249  
 FT DOMAIN 579 2250  
 FT DOMAIN 580 2251  
 FT DOMAIN 581 2252  
 FT DOMAIN 582 2253  
 FT DOMAIN 583 2254  
 FT DOMAIN 584 2255  
 FT DOMAIN 585 2256  
 FT DOMAIN 586 2257  
 FT DOMAIN 587 2258  
 FT DOMAIN 588 2259  
 FT DOMAIN 589 2260  
 FT DOMAIN 590 2261  
 FT DOMAIN 591 2262  
 FT DOMAIN 592 2263  
 FT DOMAIN 593 2264  
 FT DOMAIN 594 2265  
 FT DOMAIN 595 2266  
 FT DOMAIN 596 2267  
 FT DOMAIN 597 2268  
 FT DOMAIN 598 2269  
 FT DOMAIN 599 2270  
 FT DOMAIN 600 2271  
 FT DOMAIN 601 2272  
 FT DOMAIN 602 2273  
 FT DOMAIN 603 2274  
 FT DOMAIN 604 2275  
 FT DOMAIN 605 2276  
 FT DOMAIN 606 2277  
 FT DOMAIN 607 2278  
 FT DOMAIN 608 2279  
 FT DOMAIN 609 2280  
 FT DOMAIN 610 2281  
 FT DOMAIN 611 2282  
 FT DOMAIN 612 2283  
 FT DOMAIN 613 2284  
 FT DOMAIN 614 2285  
 FT DOMAIN 615 2286  
 FT DOMAIN 616 2287  
 FT DOMAIN 617 2288  
 FT DOMAIN 618 2289  
 FT DOMAIN 619 2290  
 FT DOMAIN 620 2291  
 FT DOMAIN 621 2292  
 FT DOMAIN 622 2293  
 FT DOMAIN 623 2294  
 FT DOMAIN 624 2295  
 FT DOMAIN 625 2296  
 FT DOMAIN 626 2297  
 FT DOMAIN 627 2298  
 FT DOMAIN 628 2299  
 FT DOMAIN 629 2300  
 FT DOMAIN 630 2301  
 FT DOMAIN 631 2302  
 FT DOMAIN 632 2303  
 FT DOMAIN 633 2304  
 FT DOMAIN 634 2305  
 FT DOMAIN 635 2306  
 FT DOMAIN 636 2307  
 FT DOMAIN 637 2308  
 FT DOMAIN 638 2309  
 FT DOMAIN 639 2310  
 FT DOMAIN 640 2311  
 FT DOMAIN 641 2312  
 FT DOMAIN 642 2313  
 FT DOMAIN 643 2314  
 FT DOMAIN 644 2315  
 FT DOMAIN 645 2316  
 FT DOMAIN 646 2317  
 FT DOMAIN 647 2318  
 FT DOMAIN 648 2319  
 FT DOMAIN 649 2320  
 FT DOMAIN 650 2321  
 FT DOMAIN 651 2322  
 FT DOMAIN 652 2323  
 FT DOMAIN 653 2324  
 FT DOMAIN 654 2325  
 FT DOMAIN 655 2326  
 FT DOMAIN 656 2327  
 FT DOMAIN 657 2328  
 FT DOMAIN 658 2329  
 FT DOMAIN 659 2330  
 FT DOMAIN 660 2331  
 FT DOMAIN 661 2332  
 FT DOMAIN 662



QY 293 SLFT----KVLLOMLQWLDSPEVGGPLRAQLRMLASQASAGR----- 331  
 Db 324 DLEIDMAELMAGLMDGHRGALGGPGKALGSEGDADPGKALRGSEFPDDGVDD 383  
 QY 332 -----RLSDVRGGLL-----RLAE-ALAFRODLEVVSTVRVITLNSGECC----- 373  
 Db 384 DRLVOEVRRLSATLGGLODHDGSRSLSPGALFPKLMKRRKSSERPEASLSSEETACVE 443  
 QY 374 SVEPDLISKVLQGLIEVRSPLHELLTFAFSATADAAAPACKRVVYVSSLLQEEBPL 433  
 Db 444 NVKSQTSKDLG-----QPHSEPGAGFGLQNMOPG-----SEEDSL 485  
 QY 434 AGKRGADGSGLEAVRLGPSS-----GLVYDLEMLDPEVSSCPDLQLRLFSRRKGQ 489  
 Db 486 PAGQGEALGDGLQ-LEVAKPSSEEARGYVTDRDPLRPEEGQVVEDVARLL----- 535  
 QY 490 AOVPS 494  
 Db 536 -QMP5 539

## RESULT 14

GEM4\_HUMAN STANDARD: PRT; 1058 AA.

AC P57678; Q9NZ57;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Component of gems 4 (Gemin4) (p97).  
 GN GEMIN4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND MASS SPECTROMETRY.  
 RX MEDLINE=20191885; PubMed=10725331;  
 RA Charroux B., Pellizzoni L., Perkinson R.A., Yong J., Shevchenko A.,  
 Mann M., Dreyfuss G.;  
 RT "Gemin4: a novel component of the SMN complex that is found in both  
 RL J. Cell Biol. 148:1177-1186(2000).  
 CC -1- FUNCTION: THE SMN COMPLEX PLAYS AN ESSENTIAL ROLE IN SPLICEOSOMAL  
 CC SNRNP ASSEMBLY IN THE CYTOPLASM, AND IS REQUIRED FOR PRE-MRNA  
 CC SPLICING IN THE NUCLEUS. GEMIN4 COULD SERVE AS A COFACTOR OF  
 CC GEMIN3.  
 CC -1- SUBUNIT: FORMS A STABLE HETEROMERIC COMPLEX WITH SURVIVAL OF MOTOR  
 CC NEURON PROTEIN (SMN), GEMIN2 AND GEMIN3. INTERACTS DIRECTLY WITH  
 CC GEMIN3 AND WITH SEVERAL SNRNP SM CORE PROTEINS, INCLUDING B/B',  
 CC DI-D3, AND E.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR. LOCALIZED IN  
 CC SUBNUCLEAR STRUCTURES NEXT TO COILED BODIES, CALLED GEMS, WHICH  
 CC ARE HIGHLY ENRICHED IN SPLICEOSOMAL SNRNPs AND IN THE NUCLEOLUS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AF173856; AAF35283.1;  
 DR Genbank: HGNC:15717; GEMIN4.  
 DR MIM: 606969;  
 KW mRNA processing; Spliceosome; Nuclear protein.  
 FT DOMAIN 714 735 LEUCINE-ZIPPER (POTENTIAL).  
 SQ SEQUENCE 1058 AA; 119989 MW; 5EE1FE2EE3538D6B CRC64;

Query Match 2.3%; Score 121.5; DB 1; Length 1058;  
 Best Local Similarity 18.3%; Pred. No. 2.1;  
 Matches 216; Conservative 149; Mismatches 359; Indels 457; Gaps 52;

QY 3 ILVHNAVILLTLPGRADSEFOALLDIWPEEKPFLTAFLVDTSEBALLPDWLKLRM 62  
 Db 70 LIIIMAKV--LQPHVPTSDIETRMQEDLDFSVGMIIPTI-----NHTLTFELKSL 120  
 QY 63 IRSEVLRIVDAALQDLPEQQLLEVQSGFIPVSSKSLLOFLD--QAVAH-----DPOT 114  
 Db 121 ASGLFIQLMLLPTTICHAELERFLEHVTDTSA-EDVAFPDIMWEVWKHKHGDPELL 179  
 QY 115 LEQNTMDKNYMAHLYVEVQHERGASGGQTFHLLTSLPRRDSAPKPKSPQIQIG 174  
 Db 180 SPSAMAKHYLPALDEPH-----PPKRLRSD--PDACPIMPLIAM 218  
 QY 175 RIRVGTOL--RYLQPE-----DLAGMFIQIFPLS-PDPR-----WQ 208  
 Db 219 LLRGITQIOSRLIGRGKRCALANLADN-LYFALTEDDPQEVSAITYLDKATVTSWN 277  
 QY 209 SSSPPRVALLQALQGLQELARYVGGSPPEGTYTVRVALATLSSPGALVMSHRSH 268  
 Db 278 SDTONPYH--QALAE--KVKEARDVS-----LTSKLKPS-----ETI 313  
 QY 269 FLACPLRLQLOCOYKCYVODTGFSFLKVLLOMLQWLDSPEVGGPLRAQLRMLASQAS 328  
 Db 314 FVGCFFLHLL-----LKEW-----GEELOAVLRSSQGTSTY 343  
 QY 329 AGRLRSD-----VRGLRLAEALAFRODLEVS-----STV--RAV----- 363  
 Db 344 DSYRLCDSLTSTSQNATLYLNTSLSKEDROYSLACVDFLKTSTYVKNRLIEDIT 403  
 QY 364 -----IATLRSGQCSVEPDLISKVLQGLIE 389  
 Db 404 ASIAMAIVIQKMDRMHEVCYIFASEKKAWSDEWACAGLSNRLPREPDLVRLLETVID 463  
 QY 390 VRSPHLEELLTFAFSATADAAAPACKRVVYVSSLLQEEBPLAGGPGADGSLAEV 449  
 Db 464 V-----STDRAIPESQIRQV--HLIECTADLS-LPGKN----- 496  
 QY 450 LGPSSGLVDMLEMLDPEVSSCPDLQLRLFSRRKGQ--AGVSPFRPYLLTF-- 503  
 Db 497 KVLGILRSM-----GRKSGSEKLAVVEGEOEDLNTFFQNL 533  
 QY 504 THQSSWPTL-----HQCIRVLGSRDEQF-- 528  
 Db 534 TQASAEGLAKAVASVARKLYVHPETVKKMCSLAVNLTGHKFLAOLITFAPALRFEV 593  
 QY 529 -DPSASLDFLWACIHVPRIWOGRODQTPQKRRREIVL-----RYOGPELISIV-- 575  
 Db 594 QCPNSSATFWSCIK-ETVWM--KSTPKREKQLELLNCLMSPVKPGQIYVALLEPDE 650  
 QY 576 --ELILAEATRSQDGTACSLIQ-----ARPLLLSCCGGDDSVR 616  
 Db 651 VLKEFVLPFLRLDVEVDLSLRFIQTLEANNACREYWLQTCSPPLFLSIC----- 702  
 QY 617 KYTEHLSCIOQMGDSVIGRCRDLLOLYLQRPRLRPVPEVLLHSGCASSV----- 671  
 Db 703 ---QLDRFSKYWPLPKPKRC-----LSDRKDLAHIHELCEIVASNAETPSPDW 752  
 QY 672 -----CKLDGLHFRITLLADYSDSLAENRGAD 700  
 Db 753 IKLSLWLRKLEQLDWTYGLRKSFEGHFKECVATPFLTECKLSDEDTSOAHGYTAG 812  
 QY 701 ASM-----ACRKLAVAPHLRLRLPM-----IALHLGRTHLNFQEFROO 741  
 Db 813 TGLLAMECCCVSSGSIERSMLSLVVDGPNPEVRLFSKGLVALVYVMPQSPQEWRL 872  
 QY 742 NHL-----CFHLVGLLELLQPHVFRSEHOGALMCLSLTRLLVYKSSRLAFL 795  
 Db 873 HOLTRRLLEKOLLHVPYSLEYIQ-----FVPLP-----NLKFAOEL 909  
 QY 796 NKEVOFIHKYITYNAPAIISFLQKHA--PLHDSFDNSDLVWKLKSLAGSLSPSRD 851  
 Db 910 QLSVFLKTF-----QFLCSHSCRNMPLBEMNH-----VYKLKCSLT----- 948  
 QY 852 RTDRGLDE-----EGEESASGSLPLVSVSLFTPLTAAEMADYMKRLSRGQTV 900

Db 949 ---RLDVSRAIOAGCFWVGWGPEDD-----LTQEALEFVYTOVFCALHTMAMLRPEVCE 999

QY 901 DL-----LEVLSIDIDEMSRPRPELISFFPS-TNLORLMSSAEE 936

Db 1000 PLVVALFETITCETETLSKTPSVSSLLQRAHQERFLKSTAE 1040

RESULT 15

VINC\_CHICK

AC P12003; STANDARD: PRT: 1065 AA.

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Vinculin (Metavinculin).

GN VCL.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OC NCBI\_TaxID=9031;

OX [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=89042216; PubMed=3141928;

RA Coutu M.D., Craig S.W.;

RT "cDNA-derived sequence of chicken embryo vinculin.";

RL Proc. Natl. Acad. Sci. U.S.A. 85:8539-8539(1988).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RA MEDLINE=89246350; PubMed=2497736;

RA Price G.J., Jones P., Davison M.D., Patel B., Bendori R., Gelger B., Critchley D.R.;

RT "Primary sequence and domain structure of chicken vinculin.";

RL Biochem. J. 259:453-461(1989).

RN [3]

RP SEQUENCE OF 1-880 FROM N.A.

RC TISSUE=Embryo;

RA MEDLINE=88024106; PubMed=3117046;

RA Price G.J., Jones P., Davison M.D., Patel B., Eperon I.C., Critchley D.R.;

RT "Isolation and characterization of a vinculin cDNA from chick-embryo fibroblasts.";

RL Biochem. J. 245:595-603(1987).

RN [4]

RP TALIN INTERACTION DOMAIN.

RA MEDLINE=90078327; PubMed=2512301;

RA Jones P., Jackson P., Price G.J., Patel B., Ohanion V., Lear A.L., Critchley D.R.;

RT "Identification of a talin binding site in the cytoskeletal protein vinculin.";

RL J. Cell Biol. 109:2917-2927(1989).

CC -1- FUNCTION: INVOLVED IN CELL ADHESION. MAY BE INVOLVED IN THE ATTACHMENT OF THE ACTIN-BASED MICROFILAMENTS TO THE PLASMA MEMBRANE.

CC -1- SUBUNIT: EXHIBITS SELF-ASSOCIATION PROPERTIES.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF ADHESION PLAQUES.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: VINCULIN (SHOWN HERE) AND METAVINCULIN. ARE PRODUCED BY ALTERNATIVE SPLICING. METAVINCULIN DIFFERS FROM VINCULIN BY THE INSERTION OF A 68 RESIDUES DOMAIN NEAR THE C-TERMINUS.

CC -1- PTM: PHOSPHORYLATED ON SERINES, THREONINES AND TYROSINES AND ACYLATED BY MYRISTIC ACID AND/OR PALMITIC ACID.

CC -1- SIMILARITY: STRONG, TO OTHER VINCULINS AND TO ALPHA-CATENINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

DR EMBL: J04126; AAA49136.1; -

DR EMBL: Y00312; CA68412.1; -

DR PIR: A27884; A27884.

DR PIR: A29997; A29997.

DR PIR: A31346; A31346.

DR PIR: S03973; S03973.

DR InterPro: IPR001043; Vinculin/catenin.

DR InterPro: IPR000633; Vinculin\_2.

DR Pfam: PF01044; Vinculin; 1.

DR PRINTS: PRO0806; VINCULIN.

DR ProDom: PD025583; Vinculin\_2; 3.

DR PROSITE: PS00663; VINCULIN\_1; 1.

DR PROSITE: PS00664; VINCULIN\_2; 3.

KW Cell adhesion; Actin-binding; Cytoskeleton; Structural protein; Phosphorylation; Repeat; Alternative splicing; Lipoprotein.

KW INIT MET 0

FT DOMAIN 167 207 TALIN-INTERACTION.

FT DOMAIN 258 588 3 X 112 AA TANDEN REPEATS.

FT REPEAT 258 368 1.

FT REPEAT 369 478 2.

FT REPEAT 479 588 3.

FT DOMAIN 836 877 PRO-RICH.

FT MOD\_RES 536 536 PHOSPHORYLATION (POTENTIAL).

FT MOD\_RES 821 821 PHOSPHORYLATION (POTENTIAL).

FT CONFLICT 441 446 TAKLSD -> QLSQI (IN REF. 2 AND 3).

FT CONFLICT 700 700 Q -> K (IN REF. 2 AND 3).

FT CONFLICT 879 879 E -> K (IN REF. 3).

SO SEQUENCE 1065 AA; 116867 MW; 8D6C887D9E71E77 CnC64;

Query Match 2.38; Score 121; DB 1; Length 1065;

Best Local Similarity 20.2%; Pred. No. 2.3; Mismatches 156; Gaps 26;

Matches 125; Conservative 101; Mismatches 156; Gaps 26;

QY 11 ILTLGPRRADDSFQALDIPWPEEKLPTAFVDTSEALLDPLKLRMTRESERL 70

Db 121 LLTF-----DEAEVRKIIVCKGILEYLVVAEYVETMEDLVYTKNLCGCM--TKAMK 173

QY 71 VDAALQDLERQO-----LLEFVSFGIPVSSMSKLLQFLDQAVAHDPQL 115

Db 174 IDERQGLTQHEHVMVLSNMNTYKELLPVLISMKLFVTKNKSQGIIEALKNRHFTV 233

QY 116 EQNIMDKNYAHLYEV---QHERGASGGQTFHSLTASLPPRDSTE-----APKPS 165

Db 234 EKMSAEINIEIRVQLTSMDEDMASKDTAMKRALLDISKMKQAGWLRDPNAP--PGD 292

QY 166 SPEDPICO-----GRIVGTQLR-VLGPEDDLACMFQIPLSPDPKMOSSPPRPA 216

Db 293 AGEQAIQIIDEAGKAGELCAGKERREILGTCTLGOMTQDLADLR--ARGQATPMAMQ 350

QY 217 LALQALGOEL-ARVVOGSPVEVGRIT-----VRVQALATLTLSPFGAL-----V 261

Db 351 KAQVSGGLDLTLAKVNAARKLEAMTNSKQATAKKIDAQNMLADPNNGSEGEHRTGI 410

QY 262 MSNHRSHFLAC-----PLRLQCOYQRCVPQDTGSSFLKVL---LQ 301

Db 411 MSEARKVAELCEEPKERRDLIRLSIGETISALTAKLSLRGRGKGSPPARALAQIATSLQ 470

QY 302 MLO-----WIDSPGVEG-GPLRQLMLASQASGR 332

Db 471 NLSKTRAVANTRPVKAAYVHLEGKIEQAOQRWIDNPVDRGVQAIRGLVAE---GRR 527

QY 333 LSDVRGGLRLALAEAFRODLEVYSSVTRAVITLRS---GQCSVEPLLSVLQGLI 388

Db 528 LANYMG-----PYRDLAKCKDRVQLAQLADLARGEGESQAAIAYNAQLDLSL 579

QY 389 EVSPPHLEELLTAFFS-ATADAASPPACKPVVAVSSLLQDEEPLA-----GKRP 438

Db 580 KDLKARQOEAHTQEVSVFSFDTTPIKLAVAVATASDPVNRREVEEERAAENFENHARL 639

QY 439 GAQGSILEAVRLDPPSSSLVLDMLDEM-----LDPEVSSCPDLQLRLFSRRKKGQAO 491

Wed May 14 09:27:45 2003

us-09-929-769-7.rsp

Page 17

```

Db      640 GATLEKAAAV--GIANKTVEGIGATYKSABELTPROVNSAA-----RILL--RNPGRQAA 650
QY      492 VPSRPRIILLLETHQSSW 509      :
      |      | : : |
Db      691 YEHEF-----TMMKNQW 701

```

```
Search completed: May 7, 2003, 19:41:31
Job time : 38 secs
```

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:40:03 ; Search time: 19 Seconds  
(without alignments)  
1593.484 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 5241  
Sequence: 1 MHILVHAMVILLTGPRA.....MDPSAQISEALRLHMEAVM 1029

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUTS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136.5	2.6	2475	US-09-413-814-48	Sequence 48, Appl
2	133	2.5	737	US-08-188-582-16	Sequence 16, Appl
3	133	2.5	737	US-08-646-715-16	Sequence 16, Appl
4	126	2.4	878	US-08-708-541A-26	Sequence 26, Appl
5	124.5	2.4	5087	US-09-144-085-1	Sequence 1, Appl
6	124	2.4	577	US-09-413-814-4	Sequence 4, Appl
7	123.5	2.4	1455	US-08-726-012B-2	Sequence 2, Appl
8	121	2.3	1315	US-09-031-563-2	Sequence 2, Appl
9	121	2.3	1315	US-09-031-563-25	Sequence 25, Appl
10	121	2.3	1315	US-09-293-505-10	Sequence 10, Appl
11	121	2.3	1315	US-09-392-277-2	Sequence 2, Appl
12	121	2.3	1315	US-09-392-277-25	Sequence 25, Appl
13	120	2.3	3066	US-08-952-127-12	Sequence 12, Appl
14	119.5	2.3	819	US-09-651-656-15	Sequence 15, Appl
15	119.5	2.3	819	US-09-650-855-15	Sequence 15, Appl
16	116.5	2.2	3798	US-09-568-102-6	Sequence 6, Appl
17	116.5	2.2	3798	US-09-567-969-6	Sequence 6, Appl
18	116.5	2.2	3798	US-09-568-480-6	Sequence 6, Appl
19	116.5	2.2	3798	US-09-568-486-6	Sequence 6, Appl
20	116.5	2.2	3798	US-09-567-899-6	Sequence 6, Appl
21	116.5	2.2	3798	US-09-031-563-27	Sequence 27, Appl
22	115.5	2.2	1055	US-09-392-277-27	Sequence 27, Appl
23	115.5	2.2	1055	US-08-811-481-16	Sequence 16, Appl
24	115.5	2.2	1012	US-08-242-677-2	Sequence 2, Appl
25	115	2.2	1621	US-08-514-213A-2	Sequence 2, Appl
26	115	2.2	979		
27	114	2.2	4		

28	114	2.2	1208	US-09-463-702A-2	Sequence 2, Appl
29	112.5	2.1	804	US-08-909-954-2	Sequence 2, Appl
30	112	2.1	3072	US-09-413-814-93	Sequence 93, Appl
31	112	2.1	3079	US-09-413-814-80	Sequence 80, Appl
32	111	2.1	1004	US-08-916-352-2	Sequence 2, Appl
33	111	2.1	2627	US-08-751-189-3	Sequence 3, Appl
34	111	2.1	2627	US-09-060-836-3	Sequence 3, Appl
35	111	2.1	2627	US-09-184-445-3	Sequence 3, Appl
36	111	2.1	3056	US-08-508-836A-8	Sequence 8, Appl
37	111	2.1	3056	US-08-629-001A-3	Sequence 3, Appl
38	111	2.1	3056	US-08-874-266-2	Sequence 2, Appl
39	111	2.1	3056	US-08-642-274D-3	Sequence 3, Appl
40	111	2.1	3056	US-08-952-127-3	Sequence 3, Appl
41	111	2.1	3056	US-08-952-014C-3	Sequence 3, Appl
42	111	2.1	3056	US-09-360-416-2	Sequence 2, Appl
43	110.5	2.1	3057	US-09-360-416-3	Sequence 3, Appl
44	109	2.1	1745	US-09-031-485-33	Sequence 33, Appl
45	109	2.1	1745	US-08-847-429A-33	Sequence 33, Appl

#### ALIGNMENTS

```

RESULT 1
US-09-413-814-48
; Sequence 48, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Bayer, Stefan
; APPLICANT: Bioecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.1
; LENGTH: 2475
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-48

Query Match
Best Local Similarity 20.7%; Score 136.5; DB 4; Length 2475;
Matches 193; Conservative 124; Mismatches 341; Indels 275; Gaps 50;

QY 178 VGTQLRVLPEDDL-----AGMFLQTFPLSPDRMSSSRPALALQALGGLAR 229
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 785 IGPPLPAGACAEEDLLAASASAGS-----PPEPSANSAERPAOS-----SIAS 830

QY 230 VVQGSPEVGVITRVQLATLISPH---GGALVMSHRSHEFLACPLRLQLOVRCVP 286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 831 ATTPAPSHTSASV---AAATILETVRAYFGFAVSTDAFF----- 868

QY 287 OPTGFSLEFLKVLQMLQWLDSFGVE-----GGFLRAQLRMLASQASAGRRLSDVRG 339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 869 -ELGASSLDLVNIGQLS--DRIGREVPTLLLYDHTPQALALALSASAEAPPLRG 925
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 340 LRLALALAFRODLEVVSTRAVATLTLSGEGCSVEPPLISKVGLIEV--RSFHEE 397
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 926 -----HRASTSGTASASASTAPTPGDAHS--QPSFVREODIATIGMAFRGCA 975
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 398 LTAFTSATDAASPPACKPVVVSLLLOEEBPLAGKPGADGSGLEAVLGPSSGL 457
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 976 -LDAFNNLVEG-----VESITPFEDELLAAGVPREHLASTRYVR--ANGEL 1020
QY 458 VDMLEMDPEVSSCPDQLRLLEFSRRKKGQAOVPSFRYLLTLFLFHOSWPTLHCIR 517
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1021 TGAMD-PEPE-----FEYSARAAMVDPFRVF-----HECSMAL----- 1056
QY 518 VLLGKSREORFP--SASLD-FLMACIHPVRIWGRDQRTPOKREELVLR----- 565
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1057 -----EHGYDDPTPCASIGVAGVYNNHLPWMLRPLPHLEBQFALLTLTREFPAPL 1110
QY 566 -----VOGEL-----ISVELLIAEAPTSQSDTACSLIOARPLDLLSCCG- 610
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1111 LSYKVGIRGPAISLOTGACSTSLVAIGTACRELA-----GACOMALAGVTASIERGCF 1165
QY 611 -----DDESVRYKTEHLSGCIQGWGDSVLGRCRDILLLOLYLQRPRLVPEVLLH 662
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1166 HQEGYILSPGHTSRSPDAAAGTV--FGDGV-----GMVLK-----PLAQL-- 1206
QY 663 SEGAASSVYCKLDGLIHRFTLLADTSDSBALENRG-----ADASMKRKLAVAHPLLL 717
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1207 ADDDTIHAVIKGIGINNDAKRGVFTAPSRAGOTEATRALRDAGVANSRYVE----- 1261
QY 718 RHLPMIALHGRTHLNFQEFROONHLSCLFVHGLLELLQPHVRESEHOGAL--WDCLL 775
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1262 -----AHGTA-----TRMGDPIVEALTO-----AFRAEADGPIPGSCLL 1297
QY 776 SFIRLLNFKSSRHIAAFINKFQFIHKYIT-----YNAP-AAISFQKADPLHDLSFD 830
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1298 GSKSVNGHNAAGVAGVLYALLOHRLPISLYQSNPHIDFA--ASPR-VNGQ 1353
QY 831 NSDLVYMLK--LLAGLSLPSRDRDRTDGLDEGESESSAGSLP----- 872
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1354 TSDMVAPEGTRLAGVSSFIGTNAHLIYEAPKALPTAALSTEPNDLDAGDAGLV 1413
QY 873 VVSLEFPLTAEMAPRM-KLSRGQTEVDELEVSID---EKSRRL-PELLSFEFTNL 927
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1414 LPIASAPPTALAHIAINLANHLEHPTI-----ALADVALTLQGRQWHRHSLI---- 1464
QY 928 QRLMSAECCRMIAESL-ALRSMONSPSIAAFLP-----TFWV-CGSGDFEVVOT 978
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1465 -----CRNRTEIKILRAVYHSAEVPAPVSDAPRCVFLFPGGAGVPSMARD 1514
QY 979 ALRNLPEVAL--LQEHAAVL--HRAFLVG 1005
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1515 LVNCPDFALHLDPCLDQLAELLPEDPICILFG 1547

```

```

RESULT 2
US-08-188-582-16
; Sequence 16, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLHR, HOHACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ. ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-188-582-16

```

```

Query Match 2.5%; Score 133; DB 1; Length 737;
Best Local Similarity 20.8%; Pred. No. 0.00055;
Matches 135; Conservative 78; Mismatches 233; Indels 202; Gaps 28;

```

```

QY 135 RNASGGTFFHSL-----LTASLPPRRDSEAKPKKSP-----QDPIGGGIRV 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 54 KGAGAAVT--OSSLRTPTATTSGIRATLTPVLAIRLPQNPNTIONFOLPDMVLVRS 112
QY 179 GTQLRVLPEDDLAGMFLQIF-----PLSPDPMOSSSPRPVALALQALGOE-LARVQ 232
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 ENGLLMIPQALAQMOQAHAQPOQTMAPRAPPTPSAP--PVGISVQAGPTPIANQVT 171
QY 233 GSPVPGIT--VRYLQALATLSSPH-----GCAVMSHMSHFLACLLQLQLOYOR 283
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 PTTIKVQSOAQTTVPATLQSPGVQPOLVIGAA--OTASLGTAVAVQGTGPOR 226
QY 284 CVPDDTFESSL-----FLKVLLOMQWIDSPVEGGPRLAOLRMLASQASAG 330
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 227 TVGATITSSAATETMBENKCKNFSLTIK--LASSKOSTETANVKELYONLIDG 282
QY 331 R-RLSDVRGGLR-----LAEALAFRDLE----- 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 KIEAEDFTSLRYELINSSPPQVLPFLKRLPALRQLTPDSAAFIOQSQOQPPPTSQAT 342
QY 355 -----VVSSTVR-----AVATLRSGEQSY-----EPDLISKVLQ- 386
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 TALTAVALSSVORTAGKTATATYTSALQPPVLSLTQPTQVGVKQGPPTPLVIOQPPKG 402
QY 387 -----LIEVRSPHLEELITAFSAATADASPPACKPVVVSLLLODEEP 432
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 ALIRPGVITLTQTPMAVLRQPHNRIMLTTOQIQLNPLQVPPVYKRAVLDTALSAVSA 462
QY 433 LAGG-----KPGA-----DGSLEAVRLGSSGLVDWLEMDPEVSSCPD- 474
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 463 QAAAGKRLKEPGGGSEFRDDDDINDVASWAGVNLSESRILATINSELGTLTRSKDE 522
QY 475 -----LQRL-L-FSRKRGKGAQVPSFRPYLLTLFTHSSMPTLHCIRVLLGK- 522
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 TFLQAPLQRLRIIEIGKHH-----ITELHPDVAVSYSHAT-----QORLQNLVEKISFT 572
QY 523 -----SREORFDSASLDFLMACIHPVRIWGRDQRTPOKREELV-----L 584
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 AQQKNSYKDDRYEDQASDVRAQLKFF-----EQDLQIKQKKDEBEREILIRAKRSR 626
QY 565 RVQGPFLISL-----VELLIAE-AETRSQDGTACSLIOARPLDLLSC 607
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 627 ROEDPEQLRLKQKAKEMQOQELAQMRQDANLALAIIGRRKKRYVC 674

```



```

OY 17 PPRADSEFOALLIDWPEEKPLPT-----AIVDTSEAL-----LLPD 56
D 22 PTAGOVVEELLIKVWVPEDPLASPSRLAKFLRENGCYKVLORSLPENEEVETDOLLP 81
OY 57 WKLKRIKRSYLR-LVDAALQDLE-----PQOLLEFVQS 89
D 82 LAMROI EGAVLKPTLSLPGDGEYEPKYPTRHPSKEKNAPPDIALIKOMITYLELO- 140
OY 90 FGIVSMKSLQFLDOAVNHPQTEQNIMDKNY-----MAHVEVOHERGAS 138
D 141 --VPEA-----NEGKDEVTLTQNTROKAYSGITMGQANRLVMAKEVATGNPN 189
OY 139 G-----GQTFHS-----LITASLPPRODSTEAPKRSSEPDIGGRLRVGTOLRVGPEDD 190
D 190 KDKPLKGYTFESTIAQLDITLP-----VGPPGEDDKFWVPL-----TRVPSRMVL--TGD 238
OY 191 LAGMFIQIFLSPDPHMOSSPPR-VALAQAOLGOELAVNGVSPVPGITRVQALA 249
D 239 VGDGF-EVEDYLRKINKLSSGLPYGRTKGFTIGEMIA-----ISQFLRELS 286
OY 250 TLSSPHG-----GALVMSHMRSHFLACPLRLQLCO-YOR-----CVPDTGCF 291
D 287 TLKQAGTGSNKKLLSLSDYWLSCGLLPKAEKYDKSTWLTKTRNIMASPT-- 344
OY 292 .SFLKVLQMLQW---LQSP---GVEGGLRAQLRLMLASQASAGRRLSDVRGGLRLA 344
D 345 -----HMTSMITWPMVMSNPVNLNIEGCP-----SLTKFNPFRGGLNRIV 386
OY 345 E-----ALAFRODLEVSVTRAVIATLRSGEQS---VEPDLSVILGGLIEVRS 392
D 387 EMILAEERKALYADNIITYVHSNTWYSLDEGEANCTQHMQAAMYLYLTRKWSNGD 446
OY 393 PHLEELLTAFFSATADAASFPACKPVVVSLLQEEPLAGKPGADGSLAEAVRLP 452
D 447 PMNQWATFAMNIADA-----LVVDSCLIMLQIKTYGO--GSGNAATFINNH 495
OY 453 SSGLLVD-WLEMIDPEVSSCPDQLRLFLSRKKGQAQVPSFRPLTLFTHQSSWPT 511
D 496 LSTVLVDQMLMRQPR-----PDSE-----EFKSIEDKLGINFKIERSID 536
OY 512 LHQICIVLL-----GKSRQREFDPSASLDPL-WAC-----IHVRIMQGRDQT 554
D 537 IREKQLQVLLAQPGYLSGVEPEQSSPTVELDLGMSATYSKDLGIVYVL----- 588
OY 555 PQRREBELVLRVOGPELISLVELILAEETRSDQDGTACSLIQ--ARL-----PLL 605
D 589 ---DKERLECSAAYPKGVENKSL-----KSKYGLEQAYKVVRYEALRLVGWNYPLLN 638
OY 606 SCCGGDESVRYKTE---HLSCIOQWGD-SYLGRCRDLLOLYL--QRELRVPVP 657
D 639 KACKNNAGARRHLEAKGPELDEFLAEMSELSEGEAFEGFNITKLVTSSELAELNRPV 698

```

RESULT 5  
US-09-144-085-1  
Sequence 1, Application US/09144085  
Patent No. 6280999

GENERAL INFORMATION:  
APPLICANT: Gustafsson, Claes  
APPLICANT: Betlach, Mary C.  
APPLICANT: Ashley, Gary  
APPLICANT: Julien, Bryan  
APPLICANT: Ziemann, Rainer  
TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
FILE REFERENCE: 30062-20020.20  
CURRENT APPLICATION NUMBER: US/09/144, 085  
CURRENT FILING DATE: 1998-08-31  
EARLIER APPLICATION NUMBER: 09/010, 809  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1  
LENGTH: 5087  
TYPE: PR  
ORGANISM: Sorangium cellulosum  
US-09-144-085-1

Query Match 2.4%; Score 124.5; DB 4; Length 5087;  
Best Local Similarity 21.0%; Pred. No. 0.12;  
Matches 229; Conservative 122; Mismatches 392; Indels 349; Gaps 47;

```

OY 190 DLGMLFIQIFLSPDPHMOSSPPRVA--LALQALGOELAR--VVGSEVPGI--TVR 243
D 645 DYASHSAHMDLPELAGAKSSLRPKATQLPPTSTYAGEVSRGALDEWCRNLQTVR 704
OY 244 VLQALATLSSPHGALVMSHMRSHFLACPLRLQLCOYQCV---PQDTGFSSFLKVL 299
D 705 LDRALSKLLEDGHGVFEVSAH--PYLAMP LTTACGAQGVVVSGLQRDGGLSQLYRTL 762
OY 300 LQW-----LQWLSPPVYEGGLRAQLRLASQ-A 327
D 763 GOLHVOGHEVDMARVLRPHGGRVLELPTVAFORQRYLEAPRARGDVSSAGLRMAAHPPL 822
OY 328 SAGRLSDVVG---GLRLAEALAFROD-----LEVSVTRAVIATLRSG 370
D 823 GAATKLADGEGCHLFTGRSLAEHAMLRDHGVFGQVFPRTGMLEVALAARAVGS--RSL 880
OY 371 EOCSEVPEPLISKVLOGLEVRSPHLEELLTAFFSATADA--SPFPACKPV 419
D 881 SELTAEPLV-----LAEDGAARLQWIGA-----PDAAGREVGLYSQPEHAPEDAPV 929
OY 420 VVVSLLQEEPLAGKPPADGSLAEAVRLPSSGLLVWLEMDPEVSSCPDQLRL 479
D 930 VQHATGVLTDEL--GIP-----DELDELKMPVPGAEEVDLSG 966
OY 480 LFSRRKKGQAQVPSFRPLTLFTHQSS-----WPTLHQCIRV 518
D 967 FYERLRERGLHYGPTPG--LVLSRGCTRLXGRVVLPGTEKDRAEAYGLHPVLMQALQ 1024
OY 519 LKGSRQREFDPSA-SLDPLMACIHVPRIWQGRDQTPQKRREBELVLRVOGPEL----- 571
D 1025 VLGAAGEGHWEADALFMPFSMA-----DAATHATGSELRYVLELETDGSTQA 1073
OY 572 -----ISLVE----- 576
D 1074 TASICAADAAGQPVASGALRLRRVTAEOURLAVTRDQHLRYVFOVPVLAQPLEFAGS 1133
OY 577 -LTLAEETRSDQDGTACSLIQARLPLTLSCCGDESVRYKTEHLSCIOQ---WGDS 632
D 1134 LVVLGAAGEGQLADTLGATAIADLALRAMIEGAPTPIRVAVIDTMAASSPSRSDVAGSS 1193
OY 633 VLGRCHDILLQYLORPELRVPYEVYLHSEGAASSV--CKLDGLHRTITLLADTSDS 691
D 1194 HEATRQALSTLQWMLSEPR--RLDAVELVWVTRGAVSAAPDAVEDLAHGLMGLIRTARS 1251
OY 692 RALENR-----GADASMACRKLTA-VAPHLILLRLHLMIAALL----- 727
D 1252 EHPERRLRLIDVTEPYDAGLARALATAAPELALGCAVLARLVQAAAEELTRAR 1311
OY 728 -----HGRTHLNPQEFROQNHLSCEFLVILGELLELP 759
D 1312 GLDPACTVLTVGAVGGLQAVTRHLVRAHGVHVLVLSRSG-----LEAPGARLELVQ- 1363
OY 760 HVRFRSEHOGA---LMDCLIS---FIRLLNKRKSRHLAFTINKVOTIHKYI--TYN 809
D 1364 ---SLEETGAEVSWACDVSKREETARVLAGI--DAARVAVLH--LAGVYHGVIOQTOT 1418
OY 810 APAISFLQKHAD---PLHLS--FQNSDLYMLKSLLAGLSLRSD- 851
D 1419 AERLAVLAPKVDGALHLELTRELDLAFAVLFPSSAAGTIGMGQGYAANTFLDAFAA 1478
OY 852 -RTDRGLDEGEESAGSLPVSLSLTP-----LTAEMAPYMKRLSGQVVEDL 902
D 1479 HRGRGRL-----AATSL--AMGVWTPAGCGMAAQLGAELARSRGIVVMSVEEG 1527

```



Db 382 QEVHMRVLSFVSALVCFPEAOQLLEDMVRLMAOAFESCCQDSMTVAFVLVROALEG 441  
 QY 112 POTLEONIMDKNYMAHLEVOHERGASGO-----TFHSLTASLPERRDSTE 159  
 Db 442 PSA-----FLSYADMWRKASGSTRGYHGCSKALVLFPLTSLVLPFESRIVLOVH 492  
 QY 160 APPKSSPEQ-----PIQGRIRVG---TOLRVLGPEDDLGMFLQIFPLSPDRMOS 209  
 Db 493 ILHPPLVPSKYRSLTLDYISLAKTRLDLVKSTIENMGLYEDL----- 534  
 QY 210 SSFPRVALAQAALGOELAVYOGSPFVPGITVAVLQALATLLSSPHGALVMSHRSHF 269  
 Db 535 SSAGDITEPHSOAL-QDVEKAIMVEHTGNIPYVME--ASIFRRPY-----YVSHF 583  
 QY 270 LACPLLRLQLOYORCVPODTGFSSLFKLVY-----LOMLQWLDSPGVEGSPLEPAOL 320  
 Db 584 LPAULT-----PRVLPKVPDSRVAFIESLRADKIPPSLYSTVCOQSAAEKPEPAAL 637  
 QY 321 RMLASQASAGRRLSDVAGGLRLAELAFRODLEVSVSTYRAVATYTR-----SGECCSV 375  
 Db 638 GVRAPENSAEERPLGOLTAALGELRASMTPDSORVDVISAQVAVISERLRAVLGNHEDDSV 697  
 QY 376 EPDLISKVLOGLEIVRSPLHEE-----LTAFFSATADAASPPF----- 414  
 Db 698 E-----ISKI-----QLSINPRLEPREHINAVDLLITSFCOMLMAASVAPREKGPMALEVR 751  
 QY 415 -ACKPVV--VVSSL--LLOEERPLAGKPGADGSLAVRLGPSSGLLVMDLEMLDEVV 469  
 Db 752 TWCGRVPAVLTRCLQRLRHQGFSLSA-PHYGLAALAVLHIGESRSAL-----BEVD 802  
 QY 470 SSCPDQLRLLFSSRRKKGQAQVPSFRPYLLTLFTHOSSMTLHQC---TRVLLGKRQ 526  
 Db 803 VGPR-----APAG--LPVPALFDSLTCRTDSEFLCKFCYAAISYSCLKRESSQ 851  
 QY 527 RFPSPAS-----LDPL-----MACIHP-----RIV 547  
 Db 852 SROTLCCLSPGLIKKQFLMFRLFSARQPLSEEDVASLSMRPLHLPDSADWQMAALSIW 911  
 QY 548 QGRDQTPQKRRE-----ELVLRVQGBELISLV-----ELIAEAETR 585  
 Db 912 THRFREVLKEEDVHLTYODMLHLEIO--PEADALSDTERODFHQWAIHNFHPRESSAS 970  
 QY 586 SQ--DGDVAAASLQARLPLLLSCCCGDESVKYTEHLSCIOQWGSVLC----- 635  
 Db 971 GCGDGD-----LQACTITVNALMDFHQSRT-DH-----SENSDLVFCGRTGNDII 1018  
 QY 636 RCRDILLQYLQRPBELRVPEVILHSEGAASSVCKLDGLIHRFTLLADTSDSRAL 695  
 Db 1019 SRLQEMVADLEIQ-DLIVPL-----GHTPSQEHFLEIFRRRLQALT-----S 1061  
 QY 696 NRCGADSMACRKLAVAHPLLLRLHPLMIALHGRTHLNFOFRQONHLSG---PLHVL 751  
 Db 1062 GWSVAASLQORELIMYKRILR-LP--SSVLCG-----SSFOAEOPITARCEQPFHLY 1112  
 QY 752 GLELLOPHVFRSE-----HOGAL-WDCLLSFRLILNKRKSRHLAFAINKRVQFIHK 804  
 Db 1113 -----NSEMRFCGHGALITODITAHFFRGLNACLKSRDPSL---WVDTLA 1157  
 QY 805 YITYNAPAAISFLQKHAADPLHLSFDNSDLVMLKSLAGL-----SLPSRDRTD 854  
 Db 1158 KCGTKCPLLIT-----SALVWMSPLEPVLLCRMRRCOSLPRELQKL- 1200  
 QY 855 RGLDEGEESBSAGSLPLVSVSLFTPLTAEMAPYMKRLSRGO--TVEDLLE--VLSDI 909  
 Db 1201 -----OEGRQFAP-----FLSPEASAPAPNPMDLSAALHFAIQOVRREENIKROL 1246  
 QY 910 DEMSRRPPELISF-FSTNLORLMSSAECCRNALFSLALSMQNSPS---IAAFLETF- 964  
 Db 1247 KKLDCEREELLVLFELFSLMGLSS-----HLTNSSTDDLPAFHVCAAILCILE 1296  
 QY 965 -----MYCLGSDPEEVQGTALRNLP-----YALLCQDEHA--VLLCHRAFL- 1003  
 Db 1297 KKRISWALFQLTESDLRLRLRLRVAPDOHTRLLPAPYSLSYFHEDAAIREEAFLHV 1356

QY 1004 -VGMY 1007  
 Db 1357 AVDMY 1361  
 RESULT 8  
 US-09-031-563-2  
 ; Sequence 2, Application US/09031563A  
 ; Patent No. 6022708  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Frederic de Sauvage  
 ; APPLICANT: Arnon Roshenthal  
 ; TITLE OF INVENTION: Fused  
 ; FILE REFERENCE: P1272  
 ; CURRENT APPLICATION NUMBER: US/09/031,563A  
 ; CURRENT FILING DATE: 1998-02-26  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SEQ ID NO 2  
 ; LENGTH: 1315  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-031-563-2  
 Query Match 2.38: Score 121, DB 3: Length 1315;  
 Best Local Similarity 20.08: Pred. No. 0.025;  
 Matches 243: Conservative 130: Mismatches 426: Indels 414: Gaps 54;  
 QY 44 LVOTSEBALLPMLKRMIRSEVRLVDAADLPQQLLTFVQSGIPVSS-MSKILQ 102  
 Db 235 LTKPRRRLSPDLLYHPFLAGHYTITTEPAGPDL-----CTPFTSRLLPELO 282  
 QY 103 FLDAQVAH--DPQTLQONIMDKNYMAHLEVOHERGASGQTFH----- 144  
 Db 283 VLKDEQAHLRLAPKGNQSRILITQAYKRMAEAMQKHONTGPALQEDKTSKVAFGTAPLP 342  
 QY 145 -----SLT-----TASLP--PRDSIEAKPKPSPEQPIQGRIR 177  
 Db 343 RLGATPQESSLILAGILASELKSSWAKSGTGEVPSAPRENRTTPDCERAFPE----- 394  
 QY 178 VGTOLRVLPED-DLACMFQIFPLSPDRMQ-----SSSPRPVAL----- 217  
 Db 395 ---RPEVLQGRSTDVVD--LENERPDSDNEMQHILLETFEPIDLKAPLTLICNDFCOR 449  
 QY 218 ---ALQOALQOELARVVOGSEVPVGTIVRVLQALATLLSSPHGALVMSHRSHFLACPL 274  
 Db 450 IQSOLHEAGGQILKIGILEGASHI---LPAFRVLSLSSCSDSVALYSFCREAGLPLGL 505  
 QY 275 LRQL--GOYQRCVPODTGFSSLFKVLLOMLQWLDSPGVEGSPRLAQLRMLASQAS---- 328  
 Db 506 LSLRHSQESNSLQOQSMYGT-FLQDLMVAIQ-----AFACFTFNERSQTSLSQ 555  
 QY 329 ---AGRRLLSDVRGL-----LRLAELAFRODLEVSVSTVAV-----TATLRSG 370  
 Db 556 VFOGAANLFLDLQKLAQPDSDSQTLRBSIMCFYLCAMQNSAISAAPFSSLLTT 615  
 QY 371 ECGSVF-----PDLISKVLOGLEIVRSPLHEE-----LTAFFSATADAASPPAC 416  
 Db 616 QQVVLDELHGLTVPQPLRVHTPGAPQVSOPLRQSESDIPGALISSALAAICTAPVGLPDC 675  
 QY 417 -----KPY-----VYVVSILLQ-----EE 431  
 Db 676 WDAKEQVCWHLANO LTEDSSQLRPSLISGLQHPILCHLKLKLVYSCLVSEBGLCRLLQOE 735  
 QY 432 PLAGGKRGADGGSLEAV-RLGPSGLLVMDLEMPVSSCPDLQRLRLFRRKRG--G 488  
 Db 736 PLA-----LESFLMLQGVKVVWDVESEVTLYF-----LSLVFLRLQNLPCG 779  
 QY 489 QAQVPSFRPYLLTLFTHOSSMTLHQCIRVLLGKSRQ----- 526  
 Db 780 MEKIGS--DVATLFTF-SHVSVLSVAACHLQGLGGQVTFDQGPMEWMAAATHALSAP 835  
 QY 527 ---RFPDSASLDPLMACI-----HVPRIWQGRDQTPQKRREBELVLRVQGP 570

Db 836 AEVRLTPGSCGFYDGLILLQLLTLEOGKASLIDRMSSEKMTVLHMFMSVLRLL--PE 893  
Qy 571 LISLV--ELLIAEAETRSOD-----GDTAACSILQARL-----PILLSCCCGDESVRK 617  
Db 894 EASAOEGELSSPPSPPEPMTLISPQGMALLSLAMATFTQEPOLCISCSOHSTIMS 953  
Qy 618 VTEHLSGCIQOWGDSVLGRCDLLLOLYLORPELRVPPEVLLHSEGAASSVCK---- 673  
Db 954 ILKHL-----LCPSFLNQL--RQAPHSSELPVYV-----SVQQLCF 990  
Qy 674 ---LDGLHRTLLADTSDSRALENRGADASMACRLAVAPL--LLRHLEPMIALHIG 729  
Db 991 PFALDMDADLLIVLADLRDSE-----VAHLIQVCCYHLEPLMQVELP- 1033  
Qy 730 RTHLNFQERQONHLSCLFVHVLGELLEOPHFVRESEHOGALMDCLSFIRLLLNKRSR 789  
Db 1034 -----ISLITRLALMDP-----TSLNPFVNTVSASPR 1060  
Qy 790 HLAAFINKFVOFIHKYITYNAPPAISFLQKHADPLHDSFDSNDLVMLKSLAGLSPSR 849  
Db 1061 TIVSLTVAALLSDOPLTSLDLSLTAHTARVLSPEH--LSF-----IQELLAGSDESTR 1112  
Qy 850 DDRTRDGLDEGEESSEAGSLPLVSLSFTPLTAEMAPYMKRLSRGQTVEDLLEVLSDI 909  
Db 1113 PLRSLIGHPE-----NSVRAHTYRLIGHLLQHSMLRGALQOSGILLSLLGLGDK 1164  
Qy 910 DEMSRRREILSFSTNQLRLSSAECCRNIAF--SLALRSQNSPSTAAFLPFMVC 968  
Db 1165 DPVVR-----CSASFAVGNAAYQAQPLGPALAA--VPSMTOLL 1201  
Qy 969 GSODEFV---VOTALRN-----LPEYALLCOEHAVALLHRAFLVGMYGM--DPSAQIS 1017  
Db 1202 GDPQAGIRRNVASALGNIGPELGELLQCEVPRQL-----EMACGDQPNVK 1250  
Qy 1018 E---ALRIHME 1026  
Db 1251 EALIALRSLQOE 1263

RESULT 9  
US-09-031-563-25  
Sequence 25, Application US/09031563A  
Patent No. 6022708  
GENERAL INFORMATION:  
APPLICANT: Frederic de Sauvage  
APPLICANT: Arnon Koshenthai  
TITLE OF INVENTION: Fused  
FILE REFERENCE: P1272  
CURRENT APPLICATION NUMBER: US/09/031.563A  
CURRENT FILING DATE: 1998-02-26  
NUMBER OF SEQ ID NOS: 27  
SEQ ID NO 25  
LENGTH: 1315  
TYPE: PRT  
ORGANISM: artificial sequence  
FEATURE:  
OTHER INFORMATION: artificial sequence  
US-09-031-563-25

Query Match 2.3%, Score 121; DB 3; Length 1315;  
Best Local Similarity 20.0%; Pred. No. 0.025;  
Matches 243; Conservative 130; Mismatches 426; Indels 414; Gaps 54;

Qy 44 LVDTSEELLLPDMKLKLMISSEVLRVDAALODEPQOLLFVQSFQPVSS--MSKLLQ 102  
Db 235 LTKDPRQLSWPDLIRPFIFAGHTITTEPAGPDL-----GTPFTSRLLPELQ 282  
Qy 103 FLDDQAVAH--DPQTEQONIMDKNYMAHLEVEVOHERGASGQTFH----- 144  
Db 283 VLKDDQARIRLAPKQNSHLLQAYVYKMAEAMQKHNQNTGRALBEDKTSKAVPSTAPLR 342  
Qy 145 -----SLI-----TASLP--PRDSTEARPKKSSPEOPIGQGRIR 177

Db 343 RLGATPQSSSLLAGILASLKSMAKSGTGEVSAAPREKRTPDCCRAPEE----- 394  
Qy 178 VGTQRLVGLPEP-DLAGEFLQIFLSPDRMO-----SSSPRYAL----- 217  
Db 395 ---REVVLQGRSTVDV--LENEEPDSNEMOHLEETEPVYIQKAPRTLLCNPDFCOR 449  
Qy 218 ---ALQOALGOELAVNOGSPPEVPGITVAVLQALATLLSSPPGALVYMSHRSFLACL 274  
Db 450 IOSLHEAGQILKGLIBGASHI-----LPAFVLSLSSCSDSVALSFCRAGLPGL 505  
Qy 275 LRQL--COYORCVPODTGFSSFLKVLQMLQMLDPSVEGSGPLRAOLMLASQAS--- 328  
Db 506 LSLRHSGESNSLQQSNWGT--FLQDLMAVIG-----AFYACTFLENESQJSDSLQ 555  
Qy 329 ---AGRLSDVRGL-----LRLAEALARQDLEVVSTVAV--IATLNSG 370  
Db 556 VQOEANLFLDLGKILQAPDDEBQTLRRDSLMCTFVLCEAMDGNSRAISKAFYSSLT 615  
Qy 371 EGCYSVE-----PDLISKVLQGLEVRSPHLE-----LTPAFSATADAAPFPAC 416  
Db 616 QOVVLIDGLLGLTPOLPVHTPQAGAPQVSQPLREQSEDIPGAISALAICTAPVGLPDC 675  
Qy 417 -----KPV-----VVSLLQOE-----EE 431  
Db 676 WDAKEQVCMLHANQLTEDDSQQLRPSLISGLQHPILICHLKLYSCCLVSEGLCRILQGE 735  
Qy 432 PLAGKPGADGSLAV--RLGSSGLVDMLMDREYVSSCPDQLRLFRKRGK--G 488  
Db 736 PLA-----LESLFMLIOGKVVVDWESTETVTF-----LSLVERLQNLPG 779  
Qy 489 QAQVSPFRPLTFLTHOSSMPTLHQCIRVLKSGREO----- 526  
Db 780 MKKLS---DVATLTFH--SHVSVLSAACLQGLQCGQCVTTDIDPMEMAATALSNP 835  
Qy 527 ---RPDSASLDFLMACI-----HVPRIWGRDQRTQKREELVLRVQGE 570  
Db 836 AEVRLTPGSCGFYDGLILLQLLTLEOGKASLIDRMSSEKMTVLHMFMSVLRLL--PE 893  
Qy 571 LISLV--ELLIAEAETRSOD-----GDTAACSILQARL-----PILLSCCCGDESVRK 617  
Db 894 EASAOEGELSSPPSPPEPMTLISPQGMALLSLAMATFTQEPOLCISCSOHSTIMS 953  
Qy 618 VTEHLSGCIQOWGDSVLGRCDLLLOLYLORPELRVPPEVLLHSEGAASSVCK---- 673  
Db 954 ILKHL-----LCPSFLNQL--RQAPHSSELPVYV-----SVQQLCF 990  
Qy 674 ---LDGLHRTLLADTSDSRALENRGADASMACRLAVAPL--LLRHLEPMIALHIG 729  
Db 991 PFALDMDADLLIVLADLRDSE-----VAHLIQVCCYHLEPLMQVELP- 1033  
Qy 730 RTHLNFQERQONHLSCLFVHVLGELLEOPHFVRESEHOGALMDCLSFIRLLLNKRSR 789  
Db 1034 -----ISLITRLALMDP-----TSLNPFVNTVSASPR 1060  
Qy 790 HLAAFINKFVOFIHKYITYNAPPAISFLQKHADPLHDSFDSNDLVMLKSLAGLSPSR 849  
Db 1061 TIVSLTVAALLSDOPLTSLDLSLTAHTARVLSPEH--LSF-----IQELLAGSDESTR 1112  
Qy 850 DDRTRDGLDEGEESSEAGSLPLVSLSFTPLTAEMAPYMKRLSRGQTVEDLLEVLSDI 909  
Db 1113 PLRSLIGHPE-----NSVRAHTYRLIGHLLQHSMLRGALQOSGILLSLLGLGDK 1164  
Qy 910 DEMSRRREILSFSTNQLRLSSAECCRNIAF--SLALRSQNSPSTAAFLPFMVC 968  
Db 1165 DPVVR-----CSASFAVGNAAYQAQPLGPALAA--VPSMTOLL 1201  
Qy 969 GSODEFV---VOTALRN-----LPEYALLCOEHAVALLHRAFLVGMYGM--DPSAQIS 1017  
Db 1202 GDPQAGIRRNVASALGNIGPELGELLQCEVPRQL-----EMACGDQPNVK 1250  
Qy 1018 E---ALRIHME 1026





QY 145 -----SL-----TASLP--PRDSTEAPKPKSSPEOPIGGRIR 177  
Db 343 RLGATPOESSLAGILASELKSSMAKSGTGEVSAPRENRTTDPCCERAPPE----- 394  
QY 178 VGTOLRVIGED-DIAGMFQIIFPLSPDRMO-----SSSRPAL----- 217  
Db 395 ---RPEVLGGRSDVD--LENEPDSDNEMOHLLETTEVPIQLKAPLTLICNPFCOR 449  
QY 218 ---ALQALGOELARVVGSPPEVPGITVRVLOALATLTLSSPHGALVMSHRSHFLACPL 274  
Db 450 IQSOLHEAGQILKILLEGASHI-----LPAFVLSLSSCSDSVALYSCREAGLPGLL 505  
QY 275 LRQL--CQYRCVPDGTGFSLEFLKVLQMLQMLDSPGVGGLRAOLRLMAQAS----- 328  
Db 506 LSLRLHSGESNSLQOOSWYGT-FLQDLMAYIQ-----AVFACTFNERSQTSLSIQ 555  
QY 329 ---AGRLSDVVGGL-----LRLAELAFRODLEEVSSYRAV-----IATLRSG 370  
Db 556 VFOEAMNLFIDLGLKLLAQPDSEQTLRDLSCFTVLCAMGNSRAISKAFYSLLTT 615  
QY 371 EOCSE-----PDLISKVLOGLEVRSPHLE-----LTAFASTADAASPPAC 416  
Db 616 IQVVLGGLHGLTVPOLPVHTPGAPQVSOPLREQSEDIGALISSALAICTAPVGLPDC 675  
QY 417 -----KPV-----VVVSSILLQOE-----EE 431  
Db 676 WDAKEQVCWMLANQLTEDSSQLRPSLISGLQHPILCLHLKLVYSCCLVSEGCRLQOE 735  
QY 432 PLAGKPGADGSLVAV-RIGPSSGILLVDMLEMDPEVSSCCDLOLRILFSRRKG--G 488  
Db 736 PLA-----LESFPMILOGKVKVYDMEESTVTLVF-----LSLLVFLQMLPCG 779  
QY 489 QAOVPSFRPYLLTLFTHOSSWPTLHOCIRVLLGKSREO----- 526  
Db 780 MERIGS--DVATLFTN-SHIVSVLSVAACLLQOLGQGTFLQPRMEMMAATHTALAP 835  
QY 527 ---RFPSSASLDPLMACT-----HVPRTMOGDKQTPKRRRELVLRQOGE 570  
Db 836 AEVRLTPPGSCGYDGLILLQLLQTEOGKASLIRDMSSSEMWTVMHRSWMLRT--PE 893  
QY 571 LISLV--ELILAEATRSOD-----GDFAOCSLIQARL-----PILLSGCCDDSEVR 617  
Db 894 EASNOGELSLSSPSPERPWTLSPOGMALLSLAMATTOERPOLCLOLSOHSGLTMS 933  
QY 618 VTEHSGCIGOWGDSVLRRCRDLLQLYLQREPLRVYREVLHSEGAASSVCK----- 673  
Db 954 ILKHL-----LCPSFLNQL-RQADHSGEFLRVVL-----SYCQLLCE 990  
QY 674 ---LDGLIIRFTLLADTSSRALENRGADASACKLAVAPL-LILRLPMTAILHG 729  
Db 991 PFALDMDADLLIVYADLRDSE-----VAHLLQVCCYHLPLMOVELP- 1033  
QY 730 RTHLNFQEFROONHLSCLFVLGLLELLOPHVFRSEHOGALMPCLLSIFILLNYRKRSS 789  
Db 1034 -----ISLIRLALMD-----TSLNOFVNTVSASPR 1060  
QY 790 HLAATINKVQFTHKYITYNAPAAISFLOKHADPLDLSPDNSDLVMLKSLGLSLPSR 849  
Db 1061 TIVSFLSVALLSDQPLTSLSLAHATAVLSPSH-TSF-----IQGLLAGSDSESR 1112  
QY 850 DQDTGDLDEGESESSAGSLPVSVSLFPLTAAEKAPMKRLSGQYVEDLLEVLSDI 909  
Db 1113 PLUSLIGHPE-----NSVRAHYRLGLHLLQHSNALGALOSGGLSLLLGLGDK 1164  
QY 910 DEMSRRRPETLSFTSTNORLMSAECCRNLAFL-SIALSRMONSPSIAAFLPTFYCL 968  
Db 1165 DPVVR-----CSASFVAGNAAYAGPLGALAAA-VPSMTQL 1201  
QY 969 GSQDEFEV---VOTALRN-----LPEYALLCOEHAVALHRAFLVGYGM--DPSAQIS 1017  
Db 1202 GPOAGIRRVASALGNLGPGEELQCEVPORL-----EWACGDQDPNVK 1250  
QY 1018 E---ALRILHME 1026

Db 1251 EAALIALRSLOE 1263  
RESULT 12  
US-09-392-277-25  
; Sequence 25, Application US/09392277A  
; Patent No. 6451977  
; GENERAL INFORMATION:  
; APPLICANT: Frederic de Sauvage  
; APPLICANT: Arnon Rosenthal  
; TITLE OF INVENTION: Fused  
; FILE REFERENCE: P1272R1P1  
; CURRENT APPLICATION NUMBER: US/09/392,277A  
; EARLIER APPLICATION NUMBER: US 09/258,000  
; EARLIER FILING DATE: 1999-02-25  
; EARLIER APPLICATION NUMBER: US 60/076,072  
; EARLIER FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 28  
; SEQ ID NO 25  
; LENGTH: 1315  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: artificial sequence  
US-09-392-277-25  
Query Match 2.3%; Score 121; DB 4; Length 1315;  
Best Local Similarity 20.0%; Pred. No. 0.025;  
Matches 243; Conservative 130; Mismatches 426; Indels 414; Gaps 54;  
QY 44 LVOTSEBALLPWLKRLMRSEVLRVDAADLDLEPQQLLTVQSGFIPVSS-MSKLLQ 102  
Db 235 LTRDPKRLSWPDLIHNPLFAGHTTITTEPAGPDL-----GTFTSRLLPELQ 282  
QY 103 FLDQAVAH--DPTLEONIMDKNYMAHLVEVQHERGASGGQTH----- 144  
Db 283 VLADQAHRLRAPKNGNSRILTLQAYKRAEBAMQKHQNTGPALQEDDKTSKVAQTPALP 342  
QY 145 -----SL-----TASLP--PRDSTEAPKPKSSPEOPIGGRIR 177  
Db 343 RLGATPOESSLAGILASELKSSMAKSGTGEVSAPRENRTTDPCCERAPPE----- 394  
QY 178 VGTOLRVIGED-DIAGMFQIIFPLSPDRMO-----SSSRPAL----- 217  
Db 395 ---RPEVLGGRSDVD--LENEPDSDNEMOHLLETTEVPIQLKAPLTLICNPFCOR 449  
QY 218 ---ALQALGOELARVVGSPPEVPGITVRVLOALATLTLSSPHGALVMSHRSHFLACPL 274  
Db 450 IQSOLHEAGQILKILLEGASHI-----LPAFVLSLSSCSDSVALYSCREAGLPGLL 505  
QY 275 LRQL--CQYRCVPDGTGFSLEFLKVLQMLQMLDSPGVGGLRAOLRLMAQAS----- 328  
Db 506 LSLRLHSGESNSLQOOSWYGT-FLQDLMAYIQ-----AVFACTFNERSQTSLSIQ 555  
QY 329 ---AGRLSDVVGGL-----LRLAELAFRODLEEVSSYRAV-----IATLRSG 370  
Db 556 VFOEAMNLFIDLGLKLLAQPDSEQTLRDLSCFTVLCAMGNSRAISKAFYSLLTT 615  
QY 371 EOCSE-----PDLISKVLOGLEVRSPHLE-----LTAFASTADAASPPAC 416  
Db 616 IQVVLGGLHGLTVPOLPVHTPGAPQVSOPLREQSEDIGALISSALAICTAPVGLPDC 675  
QY 417 -----KPV-----VVVSSILLQOE-----EE 431  
Db 676 WDAKEQVCWMLANQLTEDSSQLRPSLISGLQHPILCLHLKLVYSCCLVSEGCRLQOE 735  
QY 432 PLAGKPGADGSLVAV-RIGPSSGILLVDMLEMDPEVSSCCDLOLRILFSRRKG--G 488  
Db 736 PLA-----LESFPMILOGKVKVYDMEESTVTLVF-----LSLLVFLQMLPCG 779  
QY 489 QAOVPSFRPYLLTLFTHOSSWPTLHOCIRVLLGKSREO----- 526

```

Db 780 MEKGS---DVALTETH-SHVVSVLSAACLGLGOGVTFDLQPMEMMAATHALSAP 835
QY 527 ---RFDPSASIDFLMACI-----HVPRIWQGRDQRTPOKREELVLRVOCPE 570
Db 836 AEVNLTPPGSGCFDGLILLLOLLEOGRKASLRDMSSSMYVLMHRSVRL--PE 893
QY 571 LISLV--ELLIAEATRSQD-----GDTAACSLIOART--PLLSCCCGDESEVR 617
Db 894 EASAOESELSSPPEPDTLLISPOGMALLSLAMATEQEPOLCISLCHGSLIMS 953
QY 618 VTEHLSGCIQOMGDSVIGRRCRDLLLOLYQREPLRVPPVYLHSEGAASSVCK--- 673
Db 954 ILKHL-----LCPSEFLNOL-RQAPGSEFLPVVVL-----SVCOLLCE 990
QY 674 ---DDGILHRTLLADTDSRALENGADASMACRLAVANPL-LLLRHLPMTAALLHG 729
Db 991 PFALMDMADLLIVLADLRSE-----VAHLQVCCYHPLMOVELP- 1033
QY 730 RTHLNEQEPKQNHLSCEFLHVLGLLELLQPHVRSHEOGALMDCLLSFIRLLNVRKSSR 789
Db 1034 -----ISLTLRLALMDP-----TSLNCFVNTVSASPR 1060
QY 790 HIAFINKEVQFIHKYTYNAPAIISFLQKHADPLHDSFNDVLMKSLLAGLSLPSR 849
Db 1061 TTVSELSVALSDOPLTSLDLSLAHTARVLSPSH-LSF-----IQELLASDESYSR 1112
QY 850 DDRTDRLDEGESESSAGSLPLVSVSLFTPLTAEMAPYKRLSRQTVDELLEVLSDI 909
Db 1113 PLRSLGHPE-----NSVRANTYRLGLHLQHSMAKRALOSQSLSLLLGLGDK 1164
QY 910 DEMSRREPELISFSTNLQRLMSSAECCRNLA-STAISMONSPTIAAFLPTMYCL 968
Db 1165 DPVVR-----CSASFVGNMAYQAGPGLPAA-VPSMQLL 1201
QY 969 GSQOFEV---VQATLRN-----LPEVALCOEHAIVLLHRAFLVGMVGM---DPSAQIS 1017
Db 1202 GDPQAGIRRVNVAALGMLGPGGLCEELLQCEVFORLL-----EMACGDPQPNVK 1250
QY 1018 E---ALRILHME 1026
Db 1251 EALIALRSLOQE 1263

RESULT 13
US-08-952-127-12
Sequence 12, Application US/08952127
Patent No. 6211336
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
APPLICANT: Tagle, Danilo A.
APPLICANT: Collins, Francis S.
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6211336Western Hwy., Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,127
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995

```

```

REFERENCE/DOCKET NUMBER: 2290.00029
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3066 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Mus musculus
US-08-952-127-12

Query Match
Best Local Similarity 18.4%; Pred. No. 0.14;
Matches 180; Conservative 136; Mismatches 254; Indels 410; Gaps 49;

QY 63 IRSEVRLVDAADL-EPQDLLEFV-----QSFQIVSSMSKLLQFL----- 104
Db 925 IRRKLLLLDSSIDLMKPLHLHMYLVLLKDLPGNEHSLPMEVDVELQPLSLVCSLHR 984
QY 105 DQVAHBPQTLQONIMDKNMAHLVEVQHERGASGGQFHSLSLTASLPRRDSTEAKPK 164
Db 985 DQVVC---KTLISNVL-----HTV----- 1000
QY 165 SSPQPIGQ-----RIRVQTLRVLPEDDLAGMFLQIFPLSPDRMOSSPREVAL 217
Db 1001 ---TNLQGSVDMESTRINAGHFLVWG-----AFHLTKKCKVF 1038
QY 218 ALQALCOELARVYQGSB-----EVPQTVRVLAOLATLLSPHAGALVMSHRSHFL 270
Db 1039 SVRMALVLCIQTLLLEADPYSEMAILNVKGDFPVNEAFSQFLADH-----HOVRML 1090
QY 271 ACPLRLQCOYQRCVPQDTGFSSFLKVLQMLQMLDSPGEGGPLRAQ--LRLMASOA 327
Db 1091 AAGSVNRFPQDMR-----QGFSSSLKAL-----BLFQOTSFNNATTTA 1130
QY 328 SAGRLSDVRGGL-----RLAEALFRODLEVYSSTV-----RAVITATLSGE 371
Db 1131 EAGIR-----GLCDSONPDLDEIYNRKSVLLMIAVLHCSPVCEKALFALCKSVK 1184
QY 372 QCSVEPDLISKVLOGLEVRPHLELLTAFFSATAAASFPKCKRVVYSSLLQEBE 431
Db 1185 ENRLEPHLVKRV-----LEKVESF-----GCR----- 1207
QY 432 PLAGCKPGADGSLFAVRLGSSGLVDMLEMLDPEVYSSCPDLQRLFLFSRRKGCQAQ 491
Db 1208 -----SLDEFMISHLDVLVLEMLNLODTER-----S 1233
QY 492 VPSFRPYLLTLEF-----HSSWPTL--HOCIRVLLKRSRQRPDPAS---LDPLMAC 540
Db 1234 LSSE-PEMLNMTSTIEDFYRSCYKILPHLYIR-----SHDEKVSIAINOIOCKWSK 1284
QY 541 IHV---PRI-----MGS-RQRTPOKREELVLRVOC----- 568
Db 1285 LLYDCEPKILVHLPRYAEETRDYSVQKR--ETATKRYDTLKGEDFLGKODOVFTSN 1342
QY 569 -PELISVELILAEATR-SODGD---TACSLIOARLPILLSCCGD-----SGDLPAHPNPYF 1388
Db 1343 LPEIV--VELLMTLHETADSADSASATRALCF-----EILSKIPDSYOKILLALCEQ 1432
QY 612 -DESVKRVTEHLSGCIQOMGDSVIGRRCRDLLLOLYQREPLRVAPVE-----VLLHSE 664
Db 1389 PSVHTQMTFVAVISNCHTKFKSIL-----EILSKIPDSYOKILLALCEQ 1432
QY 665 GAASSSVCKLDGL--HRTITLLADTSDSRALENGADASMACRLAVANPL----- 715
Db 1433 AAEFTNVEFKKRIKIVHLEFSLKLDIQS-----GIGGAMAVLVADVYITLLHYINKR 1486
QY 716 -----LRLHLPMTAALLHGRTHLNFQEPKQNHLSCEFLHVLGLLELLQPHVRSHEOG 768

```

Db 1487 SHTFVSLSRFSCLCDLLSRVCHTAVTQCK--DALESLEHV-IVGTILPLVDYOEW- 1541  
Qy 769 ALMDCLSEFIRLLNLYKRSRHLAFIN-----KFOQ 800  
Db 1542 ---EYVLLKLYVDKNDKMLSTYIKLDPFPHVIFKDLRTQOKIKYSGGFSLL 1598  
Qy 801 FHKYTYNA--PAAS-----FLOKHADPLHDL---SPDN-SDLVMLKSLAOL 844  
Db 1599 EINHTLSAVNPLPLTRLEGKLDLRQLEQHKQDMDLRRASQONPDQGIYVKLVSL 1658  
Qy 845 SLPSRDRTRDGLDEGE 864  
Db 1659 QL-----SKAVNOTGERE 1672

RESULT 14  
US-09-651-656-15  
Sequence 15, Application US/09651656  
Patent No. 6340566  
GENERAL INFORMATION:  
APPLICANT: MCCUTHEN-MALONEY, SANDRA  
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE  
TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,  
TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES  
FILE REFERENCE: IL-10689  
CURRENT APPLICATION NUMBER: US/09/651,656  
CURRENT FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 60/192,764  
PRIOR FILING DATE: 2000-03-28  
NUMBER OF SEQ ID NOS: 106  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 15  
LENGTH: 819  
TYPE: PRT  
ORGANISM: Thermus thermophilus  
US-09-651-656-15

Query Match 2.3%; Score 119.5; DB 4; Length 819;  
Best Local Similarity 22.8%; Pred. No. 0.015;  
Matches 198; Conservative 81; Mismatches 296; Indels 293; Gaps 42;  
Qy 17 PRADSEFOALLDIWPEPEKPLPAFLVDTSEALLPDMKLRINSEVLRLVDALQ 76  
Db 82 PIRADYAEIRLLKMGF---RLAVADQVEPAEEA-----EGLVREVTQL- 124  
Qy 77 DLEPOQLLFVOSFGIPVSSMSKLLQFLDOAVAHDPOTLEONIM--DKNYVAHVEVOH 133  
Db 125 -----TPGTLTQALLPREANNTLAATA----- 146  
Qy 134 ERGASGGQTFHSLTASLPERRDSTEARPKSSPEOPIGQR---IRVGTOQLRVLGPEDD 190  
Db 147 -TGDWGLAFIDVSTGEF-----KGTLLKSKSALYDELFRHRAPEVLLAPELR---ENEA 197  
Qy 191 LAGMFLQIFP--LSPDPRMOSSPPRVALLAQQALGQELARVVOGSPVPGITVRLVAL 248  
Db 198 FVAERKRREPVALSEAP--FEPOGEGPLARRAOGALLAVARATOGG---ALSVRPF-- 250  
Qy 249 ATLLSPHCGALVGMHRSHTFLACPLRLQCYORCVDPDTGFSSTFL-----KYLLOM 302  
Db 251 ---LVDP--GAFVRLPEAS-----LKALEVEEPLRGDTLFGVLDERTAPAGRRLD- 297  
Qy 303 LQWLSPGVEGGPRLAQLMLASQASAGRRLSDVGGGLRLLEALAFRODLEVSSTVA 362  
Db 298 -AMLHPLLEKPLRLARLDREVERFREGALRGVRLRLFLRLDLERLATRLLSRASPD 356  
Qy 363 VIATLRSEGGQSVEDPLISKVLQGLI--EVRSPHLEELLTAFFSATADASPFPACKPVV 420  
Db 357 LAALRRSLE-----ILPELKGILGEEVGLPDLISGLEELRAA----- 393  
Qy 421 VVSSILLQEEPLAGKRGADGSLAVARLGGSSGLVDWLEMLDPEVYSSCPDQLRL 480  
Db 394 -----LVEDPPL-----KVSEGG---LIREGYDPLD--DALRRHAEGVAYFLDLEAR-- 436

Qy 481 ESRRKGOAQVPSF-----RPY-----LITLFTPHOS-SMPTLHO 514  
Db 437 -----EKERTGPTLKGVGNVFGYYLETRRYEKKVQENRPPVOTLMDROKTYLPEKKE 491  
Qy 515 CIRVL-----LGSREORPDS-----ASDPLNACHVP----- 544  
Db 492 RERELRYLEALIKRREEEFLLRBRARKAALREARILLAEVLVYAALEVAVRHGT 551  
Qy 545 -----RIMGR--DORT--PQKREELVROGGE-----LISIV 575  
Db 552 RPFGEERLRIRAGRHDPVVERRTAEVYNDLEMAHELV--VTGNMAGKSTFLQTLAL 610  
Qy 576 ELI-----LAEAETRSODS-----DTACSLIOA--RLPLL 606  
Db 611 AQISFVAEEALPLFPGITRYRIGASDDLAKGSTFWEEMEVALLKEATERSLVL- 669  
Qy 607 CCCGDESRYKYTEHLSG--CIOQMGDSVLRGRCDLLQLYLORPELRVPE----- 658  
Db 670 -----DEVGRGTSSLDGVAIATALAEALHERRCYTLFATHYFELTALPLKMLHVA 723  
Qy 659 -----VLLHS--BGAASS--VCKLDGLIHRTFLADTSRBALENGADASMAC 705  
Db 724 KEEGGLVFFHYHQLPSPASKSYGEVYAEAGLPKEVERARALLSMAAREGALAEVLE 783  
Qy 706 RRLAV-----AHPDLLRLPMTALLHG 729  
Db 784 RLALDPRLTPELALRFLHEIKALALG 811

RESULT 15  
US-09-650-855-15  
Sequence 15, Application US/09650855  
Patent No. 6365355  
GENERAL INFORMATION:  
APPLICANT: MCCUTHEN-MALONEY, SANDRA  
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA  
TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA  
TITLE OF INVENTION: MISMATCHES  
FILE REFERENCE: IL-10284  
CURRENT APPLICATION NUMBER: US/09/650,855  
CURRENT FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 60/192,764  
PRIOR FILING DATE: 2000-03-28  
NUMBER OF SEQ ID NOS: 106  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 15  
LENGTH: 819  
TYPE: PRT  
ORGANISM: Thermus thermophilus  
US-09-650-855-15

Query Match 2.3%; Score 119.5; DB 4; Length 819;  
Best Local Similarity 22.8%; Pred. No. 0.015;  
Matches 198; Conservative 81; Mismatches 296; Indels 293; Gaps 42;  
Qy 17 PRADSEFOALLDIWPEPEKPLPAFLVDTSEALLPDMKLRINSEVLRLVDALQ 76  
Db 82 PIRADYAEIRLLKMGF---RLAVADQVEPAEEA-----EGLVREVTQL- 124  
Qy 77 DLEPOQLLFVOSFGIPVSSMSKLLQFLDOAVAHDPOTLEONIM--DKNYVAHVEVOH 133  
Db 125 -----TPGTLTQALLPREANNTLAATA----- 146  
Qy 134 ERGASGGQTFHSLTASLPERRDSTEARPKSSPEOPIGQR---IRVGTOQLRVLGPEDD 190  
Db 147 -TGDWGLAFIDVSTGEF-----KGTLLKSKSALYDELFRHRAPEVLLAPELR---ENEA 197  
Qy 191 LAGMFLQIFP--LSPDPRMOSSPPRVALLAQQALGQELARVVOGSPVPGITVRLVAL 248  
Db 198 FVAERKRREPVALSEAP--FEPOGEGPLARRAOGALLAVARATOGG---ALSVRPF-- 250



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model.

Run on: May 13, 2003, 23:29:27 ; Search time 91 seconds  
(without alignments)  
3467.806 Million cell updates/sec

Title: US-09-929-769-7

Perfect score: 5241  
Sequence: 1 MHLLVYHAWVILTLIGPPRA.....MDSQAQSEALRIHMEAVM 1029

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xjh  
-Q=/cgn2\_1/USPIO/spool/US09929769/rnmat\_07052003\_161925\_8057/app\_query.fasta.1.1223  
-DB=issued.patents\_NA -OPM=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09929769 @CGN.1.1.46 @rnmat\_07052003\_161925\_8057 -NCPU=6 -ICPU=3  
-NO.XLPHY -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCUTS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	157	3.0	33529	4	US-09-144-085-3
C 2	156.5	3.0	49377	1	US-08-764-233A-1
C 3	155.5	2.9	4080	2	US-08-446-345-35
C 4	150.5	2.9	28958	1	US-08-258-261B-6
C 5	150.5	2.9	28958	1	US-08-456-837-6
C 6	150.5	2.9	28958	1	US-08-457-342-6
C 7	150.5	2.9	28958	1	US-08-457-646A-6
C 8	150.5	2.9	28958	1	US-08-458-076A-6
C 9	150.5	2.9	28958	1	US-08-764-233A-4
C 10	150.5	2.9	28958	1	US-08-457-335A-6
C 11	150.5	2.9	28958	1	US-08-729-214-6
C 12	150.5	2.9	28958	3	US-09-028-934-6

13	147	2.8	4031	2	US-08-993-118-1	Sequence 1, Appl1
14	147	2.8	4031	3	US-08-845-528C-1	Sequence 1, Appl1
15	145	2.8	4225	3	US-08-993-118-9	Sequence 1, Appl1
16	145	2.8	4225	3	US-08-845-528C-9	Sequence 9, Appl1
17	145	2.8	4265	4	US-09-061-709-1	Sequence 1, Appl1
C 18	144.5	2.8	4154	1	US-08-131-365B-37	Sequence 37, Appl1
C 19	144.5	2.8	4154	2	US-08-668-123-37	Sequence 37, Appl1
C 20	135	2.6	8078	4	US-09-702-251-3	Sequence 3, Appl1
C 21	134.5	2.6	3879	2	US-08-916-352-1	Sequence 1, Appl1
C 22	134	2.6	5503	2	US-08-726-012B-1	Sequence 1, Appl1
C 23	133	2.5	3603	1	US-08-188-582-15	Sequence 15, Appl1
C 24	133	2.5	3603	1	US-08-646-715-15	Sequence 15, Appl1
C 25	132	2.5	3435	1	US-08-366-577-1	Sequence 1, Appl1
C 26	132	2.5	3435	1	PCT-US96-00005-1	Sequence 1, Appl1
C 27	131.5	2.5	36519	3	US-08-923-137-2	Sequence 2, Appl1
C 28	131.5	2.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
C 29	131.5	2.5	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
C 30	130	2.5	3490	2	US-08-841-483-3	Sequence 3, Appl1
C 31	130	2.5	3490	4	US-09-382-911-3	Sequence 3, Appl1
C 32	129.5	2.5	33529	4	US-09-144-085-3	Sequence 3, Appl1
C 33	129	2.5	4094	4	US-08-841-483-5	Sequence 5, Appl1
C 34	129	2.5	4094	4	US-09-382-911-5	Sequence 5, Appl1
C 35	129	2.5	7812	4	US-09-368-590-1	Sequence 1, Appl1
C 36	128.5	2.5	1371	2	US-08-910-731-1	Sequence 1, Appl1
C 37	128.5	2.5	1371	2	US-08-795-395-1	Sequence 1, Appl1
C 38	127	2.4	15378	3	US-08-785-420-1	Sequence 1, Appl1
C 39	126.5	2.4	6763	2	US-08-756-506-23	Sequence 23, Appl1
C 40	126	2.4	2827	2	US-08-708-541A-25	Sequence 25, Appl1
C 41	126	2.4	2827	4	US-08-940-968-24	Sequence 24, Appl1
C 42	126	2.4	2827	4	US-08-940-968-25	Sequence 25, Appl1
C 43	126	2.4	35081	2	US-08-752-760A-1	Sequence 1, Appl1
C 44	125.5	2.4	5228	4	US-09-428-711A-15	Sequence 15, Appl1
C 45	125	2.4	6344	4	US-08-843-417-1	Sequence 1, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-144-085-3/c  
Sequence 3, Application US/09144085  
Patent No. 6280999  
GENERAL INFORMATION:  
APPLICANT: Gustafsson, Claes  
APPLICANT: Betlach, Mary C.  
APPLICANT: Ashley, Bryan  
APPLICANT: Ziermann, Rainer  
TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
TITLE OF INVENTION: THEREFOR  
FILE REFERENCE: 30062-20020.20  
CURRENT APPLICATION NUMBER: US/09/144.085  
CURRENT FILING DATE: 1998-08-31  
EARLIER APPLICATION NUMBER: 09/010.809  
EARLIER FILING DATE: 1998-01-22  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 33529  
TYPE: DNA  
ORGANISM: Sorangium cellulosum  
US-09-144-085-3

#### Alignment Scores:

Pred. No.: 0.00168  
Score: 157.00  
Percent Similarity: 34.29%  
Best Local Similarity: 22.33%  
Query Match: 3.00%  
DB: 4  
Gaps: 41

US-09-929-769-7 (1-1029) x US-09-144-085-3 (1-33529)

Oy 150 SerleupProProAaGArGAspSerThrGluAlaProLysPProLysSerSerProGluGln 169

Db	25568	TCGGCTCCGCCCTCCCGGAGCTTATCCGGAAGAACCCAGAGGTCGGAAGCAGCCACAGCTGCT	25509
Oy	170	-----ProLieGlyGInGlyArgTLeuArgVal	178
Db	25508	CGAGCAGGAGCGGTTCCGCGACGTCAGCTTCCCTTCGGCTGGGGGTTCAGGATGTACA	25449
Oy	179	GlyThrGInLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGInLeu	198
Db	25448	GGCGCTCGACGTCGAATGGTTGGCCGGGATCGGGCGGAGCGGTATTCGCCGCTTCGGACACA	25389
Oy	199	PheProLeuSerProAspPro-----ArgTyrPInSerSerSerProArgPro	214
Db	25388	GCACCTGGCAGAAAGCTCTCAGGTGCTGCCTGCAGCCAGGCATGTGGCAGCCACCCAGCCCA	25329
Oy	215	ValAlaLeuAlaLeuGInGInAlaLeuGInGInLeuLeuAlaArgValValGInGlySer	234
Db	25328	CGATCGCTATCG-----GGTGTCTCACCCCTTCTCCGGCTCA	25290
Oy	235	ProGluValProGlyTLeuThrValArgValLeuGInAlaLeuAlaThrLeuLeuSerSer	254
Db	25289	CCTTCGCGACCCCGCGCTCCGGTCCGGCTGAGGTGAGTGCATCCAGCACCCAGCGCG	25230
Oy	285	ProHISGlyGlyAla-LeuValMetSerMetHisArgSerHisPhe-----	269
Db	25229	CCACGCTCGCTGCTCGGGGTATGCAAGCTCAGGTGCTCGGTACCGTCATCCGTCACCCGCTGC	25170
Oy	270	-LeuAlaCysProLeuLeuArgGInLeuGInGlySerGInTyrGInArgCysValProGInAspThr	289
Db	25169	GCTTCTGGACGCG---CTTCGACACTCTCACGGCCATACGGAATCCAGCCGAG----	25118
Oy	289	rglyPheSerSerLeuPheLeuLysVal-----	298
Db	25117	---GTCCAGAAAGCCACGGTCTCGGGGTCAAGCGCGCTGGTCCGTGATGCCAGTACCG	25062
Oy	289	-----LeuLeuGlnIle	302
Db	25061	CGGCGGCTCGCTCGGCCACACGACGTACGTAACCCGCTCGCGCGCTCGGCTTCGCTCAAGC	25002
Oy	302	LeuGInTyrPLeuAspSerProGlyValGInGInGlyProLeuArgAla-GInLeuArgM	322
Db	25001	CCAGCAAGGTGCTCTCAGCGCGGCTCCGCGTCCGCGCGGCGCGGCGCGCGCTC----	24948
Oy	322	eLeuAlaSerGInAlaSerAlaGlyArgArgLeuSerAspAlaArgLy-----	338
Db	24945	-----TCTCGGCCCGCGGGGCTCTCGATCCCGCTCCAGAGCGGCC	24903
Oy	339	--GlyLeuLeuArgLeuAlaGInAlaLeuAlaPheArgGInAspLeuGInValAlaSerS	358
Db	24902	TCCGCTCGCGCCGCTGAAGACAGCGCGGAGAGCGGACAGTCCAGTCCAGTGGCGATTCGCCA	24843
Oy	358	eThrValAlaArg-----AlaValIleAlaThrLeuArgSerGlyGInG	372
Db	24842	CCGAGCTCGGCCCCAGCCGACGCGACGCCCATCTCCAGCCCGCGGAGCGCTTGGGGCGAGC-	24784
Oy	372	IncysSerValGInProAspLeuIleSerLyValLeuGInGInLeuLeuGInGInValArgS	392
Db	24783	--ATACCGCGACCCCGCGGATCCGACGTGGGACTCGAGTCTGCTGTGACACATCCCTT	24726
Oy	392	eProHISLeuGInGInLeuLeuThrAlaPhe-----PheS	404
Db	24725	CTCCCGCCAGGGGCCAGTGCACACACCGTTCGACGCTGCCGCGCGCCGCGGTACC	24666
Oy	404	eAlaThrAlaAspAla-----AlaSerProPheProAlaCysLysProValAlaVal	421
Db	24665	GGCGCAGCGCGCTCCAGCCCTCGCTTCGCTGGCGCGCGTACCCAGCCTCGCGCCCGGAGCCCC	24606
Oy	421	AlValSerSerLeuLeuGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn	441
Db	24605	ACAGCCGCGGCGATGCTGCCATGAGACAGACAGCATTCACGCTCCGCTCCCGCCAGCGGTG	24546
Oy	441	spGlyGlySerLeuLeuAlaValAlaArgLeuGlyProSerSerGlyLeuLeuValAspTyrPL	461

Dd	23445	ATGCAGGTCGACAG-----	CCCCCTTCA-----	24523
Oy	461	eUGlMeLeuAaPrroGluValaValSerSerProAaPheUGlMeuAgtLeuP		481
Dd	24522	-----CCTTC-GCCGGAGCTCTGCTC-GAGCTG---CTCAAGCTCGAAC		24482
Oy	481	heSerAgtAgtLySglYtLySglGlnaGlnValProSerPheAgtPro-----T		498
Dd	24481	TCTCTCAGCGGAACAGCGGAGCTGTAGTCCGCGAGAGTGGCCAGCCGCTTCAGCGGCT		24422
Oy	488	yLLeuLeuThrLeuPheThrHisGlnSerSerTrpThrLeuHisGlnCysIleAgtV		518
Dd	24421	TCATCTCTCTCAGTCCGCGACAG-----CGCCGAGCTGTCCG		24380
Oy	518	aLLeuLeuGlyLySerAgtGlnaGlnAgtPheAaProSerAaLeuAaPheLeuT		538
Dd	24379	CGCTCCAGACAGTCAACAGCGCCAGCGT---CAACCAGATCCCGCGCCAGAAAGCTC-		24324
Oy	538	rPAaCysIleHisValProaGltIeTrpGlnGlyAgtAaPclAaTrpThrProGlnLyS		558
Dd	24323	-----GTCCCGAGCTCGCTCGCCGCGCGGCGCGCCAGCGCGCGCGCGAG		24278
Oy	558	rGAGGlnGlnLeuValLeuAgtValaGlnGlnProGluLeuLeuSerLeuValaGluLeuI		578
Dd	24277	GCCAGCAACAAGTGTCTGTGCCAGCCGCCGCGACAGCCAGCGGCAAGATGACC-----		24225
Oy	578	leLeuAlaGlnaGlnaGlnThrAgtSerGlnAaPclAaThrAaAaCysSerLeuIleG		598
Dd	24224	-----TCCGAGCCCGCCAGCGCGCGCGGATGAGACAGCTGCC-----		24186
Oy	598	lAaAaAgtLeuPProLeuLeuLeuSerCysCysCys-----GlyAaAaPAG		613
Dd	24185	-GCCGCGCTTCTCCAGCCCTCGCGCTGCCCTGCCGCGACAGCAACAATGCCCTCAG		24128
Oy	613	luseValAgtLySValThrGlnHisLeuSerGlyCysIleGlnInTrpGlyAaSerV		633
Dd	24127	AGGCCCGCGCGCGCGCGAGCCGCGACCTGATCTTGCTTC-----G		24086
Oy	633	aLLeuGlyAgtAgtCysAaAaAaPheLeuLeuGlnLeuThyTrLeuGln---AgtProGln		652
Dd	24085	TCCGACACAAGGTGTGTCAAGATGCTGTCTGCATACAGCCCGCGCTCCAGCTCCCGGACAG		24026
Oy	652	eAgtAgtValProaGltProGluVal---LeuLeuHisSerGlnGlyAaAaAaSerSerV		671
Dd	24025	TCCAGACAGCCCGCCAGAGCTCGGGGTCTCCAGCCCGCAAGCCCGCTCCAGTCCCGAC		23966
Oy	671	aLySlyLeuAaAaPheLyLeuIleHisAgtPheIleThrLeuLeuAaAaAaPThrSerAaPs		691
Dd	23965	GTACAGCGCTGTCAAGGGGTGCAGCCAGCTGTGCTGGTGGTGTGTGACGCTCCCTCTCC		23906
Oy	691	eAgtAgtAaLeuGlnAaAaAaGlnGlyAaAaAaPheAaAaSerMetAaCysAaAaAaAaAa		711
Dd	23905	GTAGACAAACAGAGCGGAGC---GCCAGATCCGGAGGTCCGCGAGAGC---		23862
Oy	711	lAaHisProLeuLeuLeuLeuAgtAaGlnHisLeuProMetIleAaAaLeuLeuHisGlyAgt		731
Dd	23861	-----CTGCACACACCGCC		23849
Oy	731	hHisLeuAaAaPheGlnGlnIleAaAaAaGlnGlnGlnAaHisLeuSerCysPheLeuHisVal		751
Dd	23848	AGCGTCTG-----GTACAGGCTG		23831
Oy	751	eGlyLeuLeuGlnLeuLeuGlnProHisValPhe-----AgtSerGlnHisGlnG		768
Dd	23830	CGGGGCGGCTGCTGTCTCCAGAGCGCTTCTGGCTCGAGGCGGTGAAGCGGCAATGAG		23771
Oy	768	lyAaAaLeuTrAaPrcysLeuLeuSerPheIleAgtLeuLeu-LeuAaAaGlyAaLySer		787
Dd	23770	CCAGCAGCTGCGCTCTCTCTTGTCAAGGCGAGCTGTGCTGTCTCAGCAGCAGCTCGAGC		23711
Oy	788	---SerAaGlnHisLeuAa		806
Dd	23710	TGCGTCTCTCCAGCGCGCGAGTGTGCAG-----ATC		23678

```

QY 807 ThrTyrAsnAlaProAlaAlaIleSerPheLeuGlnLysHisAlaAspProLeuHisasp 826
Db 23677 ACC---TCGGCCCCCGCGCGC----- 23680
QY 827 LeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeuSerLeu 846
Db 23660 ----- 23660
QY 847 ProSerArgAspArgThrAspArgGlyLeuAspGlyGluGlyGluGluSerSer 866
Db 23659 CCGAGCCGACACACGACGACCTCCGCAAGCTCTCCGGCAAGGAGGTGACACCGACGAGC 23600
QY 867 AlaGlySerLeuProLeuValSer----- 874
Db 23599 -----CAGGTGCCCCCTTACATCCGAGTACCCGTCACACGCGCTCGGACGGTTTGCAC 23546
QY 875 ValSerLeuPheThrProLeuThrAlaAlaGluMetAlaProTyrMetLysArgLeuSer 894
Db 23545 GCTTCCTCATACAGCAGCAGCTGCGCCGCTTCCTGCTCTCTGCGGCCA-CGC 23487
QY 895 ArgGlyGlnThrValGluAspLeuLeuGluValLeuSerAspIleAspGluMetSerArg 914
Db 23486 CGC-----GAGGTACGCGAG 23472
QY 915 ArgArgProGluIleLeuSerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAla 934
Db 23471 CAGCGCGCCGACGCGCGCGCGAGCTGTCAGGACGAGCTGCAGCAGCTC----- 23444
QY 935 GluGluCysArgAsnLeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerPro 954
Db 23423 -----CGCAC----GCCATCTCCGCCGCCCTCCAGCGCGC----- 23391
QY 955 SerIleAlaAlaAlaPheLeuProThrPheMetTyrCysLeuGlySerGlnAspPheGlu 974
Db 23390 -----ATTCCACAGCGCTG-----CTCCGCGCTGAGACATTTCAC 23355
QY 975 ValValGlnThrAlaLeuArgAsnLeuProGluTyrAlaLeu 988
Db 23354 CTCGGTCGCGCGCTCTTCGGGCTTCACAGCCAGTAGGCGCTG 23313

RESULT 2
: Sequence 1, Application US/08764233A
: Patent No. 5716849
: GENERAL INFORMATION:
: APPLICANT: Ligon, James M.
: APPLICANT: Schupp, Thomas
: APPLICANT: Beck, James J.
: APPLICANT: Hill, Dwight S.
: APPLICANT: Neff, Suzanna
: APPLICANT: Ryals, John A.
: TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 520 White Plains Road, P.O. Box 2005
: CITY: Tarrytown
: STATE: NY
: COUNTRY: USA
: ZIP: 10591
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/764,233A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/729,214
: FILING DATE: 09-OCT-1996

```

```

PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/258,261
: FILING DATE: 08-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Weigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: 1506/CIP6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 49377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Sorangium cellulosum
: IMMEDIATE SOURCE:
: CLONE: p98/1, pL3, and pVKM15
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 383..760
: OTHER INFORMATION: /product= "SorR"
: OTHER INFORMATION: /note= "This gene encodes a protein that is highly homolog
: OTHER INFORMATION: the reductase domains of type I PKs such as eryA from
: OTHER INFORMATION: Saccharopolyspora erythraea."
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 927..19874
: OTHER INFORMATION: /product= "SorA"
: OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs t
: OTHER INFORMATION: are known to be involved in the synthesis of polyketide
: OTHER INFORMATION: compounds."
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 942..7115
: OTHER INFORMATION: /product= "Module 1 of SorA"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 7203..12884
: OTHER INFORMATION: /product= "Module 2 of SorA"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 13455..19616
: OTHER INFORMATION: /product= "Module 3 of SorA"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 19871..46318
: OTHER INFORMATION: /product= "SorB"
: OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs ge
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 19870..24556
: OTHER INFORMATION: /product= "Module 1 of SorB"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 24638..30820
: OTHER INFORMATION: /product= "Module 2 of SorB"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 30881..35446
: OTHER INFORMATION: /product= "Module 3 of SorB"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 35528..40114
: OTHER INFORMATION: /product= "Module 4 of SorB"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 40190..46318
: OTHER INFORMATION: /product= "Module 5 of SorB"
: FEATURE:

```

NAME/KEY: misc\_feature  
 LOCATION: 46851..47891  
 OTHER INFORMATION: /product= "Sorm"  
 OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly homologous to the methyltransferase from Streptomyces  
 OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the OTHER INFORMATION: polyketide rapamycin."

## Alignment Scores:

Pred. No.:	0.00387	Length:	49377
Score:	156.50	Matches:	274
Percent Similarity:	35.038	Conservative:	119
Best Local Similarity:	24.428	Mismatches:	435
Query Match:	2.99%	Indels:	301
		Gaps:	61

US-09-929-769-7 (1-1029) x US-08-764-233A-1 (1-49377)

QY 14 ThrLeuGlyProProAlaAlaAspSerGluPheGlnAlaLeuLeuAsp----- 30  
 DB 42525 ACCTCCGGCAACCGCTCTGTC-TCGAGCGGACGAGCGGCTCCGCGATGGCAT 42583  
 QY 31 IleTrpPheProGluGluysProLeuPro-----ThrAlaPheLeuValAspThrSer 48  
 DB 42584 CGCTTCTGTCGAGGTACACCCCATCCGTCACGCTCGCCCTCCGCGAGACTGC 42643  
 QY 49 GluGlnAlaLeuLeuLeuProAspTrpLeu---LysLeuArgMetIleArgSerGluVal 67  
 DB 42644 GACGGCTACACCGCTGATCCCGTCGTCGTCGATTCGACGCGACGAGCGACACTTC 42703  
 QY 68 LeuArgLeuValAlaAlaLeuGlnAspLeuGluProGlnGlnLeuLeuPheVal 87  
 DB 42704 GCCCGCTGCTC---CTCTCTGCGGAGCTCTTACCCGAGCGCTCGCGCTGACTGG 42760  
 QY 88 GluSerPheGlyIleProValAspSerMetSerLysLeuLeuGlnPheLeuAspGlnAla 107  
 DB 42761 AACGCTTCTGCGGCTTCTGCTCCCGCAAGTCTCTCCCGACACTAC----- 42811  
 QY 108 ValAlaAlaHisAspProGlnThrLeuGlnGlnAsnIleMetAspLysAsnTrpMetAlaHis 127  
 DB 42812 -----CCCTTCCAAACGAGCGCTTCTGCTGTCACGCGCTCC---ACGCGCGAC 42856  
 QY 128 LeuValGluValGlnHisGluArgGlyAlaSerGlyGlyGlnThrPhe-----HisSer 145  
 DB 42857 GCTGCGGAGCTC-----GCCCTCGCAGGCTTACCTCGCGCGACACCGC 42901  
 QY 146 LeuLeuThrAlaSerLeuProProArgArgAsp----- 156  
 DB 42902 CTGCTGCGGCGCGCGCTGC-CCTCGCGACCGCGATGCTTGTCTTTCACAGAGCGCT 42960  
 QY 157 -----SerThr 158  
 DB 42961 CTCCTCGCAGACACCGGCTCGAAGACACAGTCGTTCGGCATACCTGCTCTCG 43020  
 QY 159 GluAlaProLysProLysSerSerProGlu---GlnProIleGlyGlnGlyArgIleArg 177  
 DB 43021 CAGCGCCCGCTCTCTGAGCTCGCCTGCATGTCCCATCTGCTGCGCTCGACACCGT 43080  
 QY 178 ValGlyThrGlnLeuArgValLeuGlyProGluAspLeuValGlyMetPheLeuGln 197  
 DB 43081 CGAAGACGTACCGCTCGA----- 43098  
 QY 198 IlePheProLeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeu 217  
 DB 43099 -----CCCCCGCTCGCTCTCCATCGACGAGGCGCGCTCTCTCTCA--- 43140  
 QY 218 AlaLeuGlnGlnAlaLeuGlyGlnGlnLeuAlaArgValValGlnGly---SerProGlu 226  
 DB 43141 GATCTCTGCGGCGCGCGAGCGGTG-CTGGAC-----GAAAGCGCGCTCTCGCTTC 43190  
 QY 237 ValProGlyIlePheValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHis 256  
 DB ----- 44117

DB 43191 ATAGCGGCGCGACGACCGCTTC-----AGATGAGCCCTTGACTCGCCACGCCA--- 43241  
 QY 257 GlyGlyAlaLeuValMetSerMetHisArgSerHisPheLeuAlaCys----- 272  
 DB 43242 ---CGCGCTCTCTCGCGAAGCTAGCCGTC-CATTGCTTGATGCTCGCGAATGCG 43297  
 QY 273 ---ProLeuLeuArgGlnLeuCysGlnTrpGlnArgCysValProGlnAspThr---Gly 290  
 DB 43298 CCCCC-----TCGCGCGCCACCGAGTGACACCGCAAGGT 43333  
 QY 291 PheSerSerLeuPheLeuValLeuLeuGlnMetLeuGlnTrpLeuAspSerProGly 310  
 DB 43334 TTCTACGACCGC-----CTCGAGACCGCTTGCG 43360  
 QY 311 ValGluGlyIlePro-----LeuArgAlaGlnLeuArgMet 322  
 DB 43361 CTGCTTATGAGCCCGAGTTCACAGGCGCTCGCGCGCTTACACAGCGCGCACGAGCT 43420  
 QY 323 LeuAlaSerGlnAlaSerAlaGlyArgArg-----LeuSerAsp 335  
 DB 43421 CTTCGCGCAAGCCAGCTCCCGGACCGCGCGAAGAGACGCGCTGTTTCCCTCA 43480  
 QY 336 ValArgGlyGlyLeuLeuArgLeuAlaGlnAlaLeuAlaPheArg----- 350  
 DB 43481 CCGCGCTGCTGACACGCGCTTGCGAGCGCTCGCTTGTAGACGACCGCAAGCG 43540  
 QY 351 ---GlnAspLeuGlnValAlaSerSerThrValArgAlaValIleAlaThrLeuArgSer 369  
 DB 43541 CTTCAGAGATCCCTTCTCGTGAGCGGAGT-----ATCGGTGCGCTCGGCTG 43588  
 QY 370 GlyGluGlnCysSerValGluProAsp-----LeuIleSerLysValIleGlnGly 386  
 DB 43589 -----GAGCACCAACCTCGCGCGCTTTCACCGCTGAGAGCG 43630  
 QY 387 LeuIleGluValArgSerProHisLeuGlnGlnLeuLeuThrAlaPhePheSerAlaThr 406  
 DB 43631 AATCTCGCGCTGCTCC-----TCTCTCGCG 43657  
 QY 407 AlaAspAlaAlaSerProPheProAlaCysLys-----ProValValValAlaSerSer 424  
 DB 43658 ACCCGAGAGCGAACCATGCGCTCGTGCAACGCTCGCGATGCGCG----- 43705  
 QY 425 LeuLeuLeuGlnGlnGluProLeuAlaGlyGlyLysProGlyAlaAspGlyLys 444  
 DB 43706 -----CCGCGTCCCGCGAGAGCTCGCAGACCGCGGAGCG-----TCCAC 43747  
 QY 445 LeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuGlnMetLeu 464  
 DB 43748 CTGATGCTCTC-----TTCGCGATGACGTGACGAGCTGCA- 43785  
 QY 465 AspProGluValAlaSer-----SerCysProAspLeuGlnLeuArgLeuLeu 480  
 DB 43786 AAGCCACCTCACCGCCATCCCGCCCGGAGCGCTCTCTCGGACACAGAG----- 43839  
 QY 481 PheSerArgArgLysGlyGlnGlnAlaGlnValProSerPheArgProGlyTrpLeuLeu 500  
 DB 43840 ---TCTGACCTGGAGACGAGGT---GCTCTGACACGCTAATACGACCTTCTGCT 43893  
 QY 501 Thr-----LeuPheThrHisGlnSerSerTrpProThrLeuHisGln 514  
 DB 43894 ACGCAGCGCGCTGACGAGGCGCTTGCCTCAAGGCTGTGATGCGCCCTCATGCG 43953  
 QY 515 CysIleArgValLeuLeuGlnGlyLysSerArgGluGln-----ArgPheAspPro 530  
 DB 43954 TCTGCGCGAAGGCGACCTTACGAGCGCGCGCGACACCGCGCTCTCGCCCT 44013  
 QY 531 SerAlaSerLeuAspPheLeuTrpAlaCysIleHisValProArgIleTrpGlnGlyArg 550  
 DB 44014 CTTCGACGCTG-----GCTCGCGGACGAGCGCTCTCTCGCGCTCGC 44061  
 QY 551 AspGlnArgTrpProGlnLysArgArgGluGlnLeuValLeuArgValGlnGlyProGlu 570  
 DB 44062 ---CTGCTGACCGCGAGCGCGCTCGCGACCGCTGAGAGAGAGCTGACAGGCGCTCG- 44117





US-08-446-345-35

Alignment Scores:

Pred. No.:	9,95e-05	length:	4080
Score:	152.50	Matches:	195
Percent Similarity:	31.83%	Conservative:	94
Best Local Similarity:	21.4e8	Mismatches:	306
Query Match:	2.91e	Indels:	314
DB:	2	Gaps:	45

US-09-929-769-7 (1-1029) X US-08-446-345-35 (1-4080)

Oy	101	LeuGIlnPheIeuAspGlnIAlaValAla-----HisAspProGlnIthr	114
Db	2952	CTTCTCTTCTCTATATGAGCAGAGCACTCGAGATAGGAGAGTCCAT-----	2905
Oy	115	LeuGIlnAsnIleMetAspIysAsnTYrMetAlaIleLeuValGIlnHisGlu	134
Db	2904	TTTAGGCGACGACATTTAAGAGACCAATTTTTTGGCATCTACATCGAG-----	2857
Oy	135	ArgGIlnAlaSerGIlyGlnIthrPheIleHisSerLeuLeuThrAlaSerLeuProArg	154
Db	2856	-----TCCTTTTCATTCCCTCTAGAGCGCGAGCCCTTCCACGA	2818
Oy	155	ArgAspSerThr-----GlnAlaProIysProIysSerSerProGIlnIProIleIy	172
Db	2817	TGTTCTTCTTCCAGAGAGAAGGTCGACACCGGCTTTTTCAGAAAGTCCCTCCGG	2758
Oy	173	GIlnGIly-----ArgIleArgValGIlyThr	180
Db	2757	CTCGGTAGCGGCTGACGTCGTGAGGTGGAGCTCGACATGGAGGGCATTCAGACCCCT	2698
Oy	181	GlnLeuArgValLeuGIlyProGIlnAspAspLeuAlaGIlyMetPheLeuGIlnIlePhePro	200
Db	2697	CTCTCCAGGCGCCGCTGGGCT-----CTCGGCTCTCGGAGCGGGGCTGTCTCA	2647
Oy	201	LeuSerProAspProArg-TyrIleSerSerSerProArgProValAlaLeuAlaLeuGI	220
Db	2646	TCATCTCTTCTTCGCGGTCTGGAC-----GTGGGCTTTGGGCTTCA	2605
Oy	220	ngln-----AlaLeuGIln--	225
Db	2604	GGATGTGCAGAGGCGCGGCGGAGGACCGAGGCGGACGAGTGGTCTCGGCGCAGGC	2545
Oy	226	-----GlnLeuAlaArgValValGIlnIlySerProGIlnValProGIlyIthrValAr	243
Db	2544	CGGAGCGAGGCTCGCGGACGTCGAGGGGGGCGCGGAGC-----CCCGCTCTCTCTCA	2488
Oy	243	gValLeuGlnAlaLeuAlaThrIleuLeuSerSerProHisGIly-----AlaIle	260
Db	2487	AGTCTCTGTCCTCTCTCTCTCTCTCTGATTAAGCATGGTGGCGTCCGACAGGAGCT	2428
Oy	260	uValMetSerMetHisArgSer-----HisPheLeuAlaIacyProIeu--	274
Db	2427	TCCTTAGGCGGTACSCCAAGCCCTCCGCGCTTCTCGGCGCTTCTGCGTGTGCTCTGG	2368
Oy	275	-----LeuArgIlnLeuCysGIlnTYrGI	282
Db	2367	TGAAAACGCTGGGCTGGAGCCACCGAAGAGGCTCGCGGCGCACCTTCGCGCGCAGC	2308
Oy	282	ArgCysValProGIlnAspThrGIlyHisSerSerIeuPheLeuIysValIleuGlnIle	302
Db	2307	C-----GGATAGGTCGCGCTCTTAGCGG-----CAAGC	2278
Oy	302	IleuGIlnItrPheuAspSerProGIly--ValGIlnGIlyIProIeuArgAlaGlnLeuAr	321
Db	2277	CTCTCAGCGCGTGGCTGAGAGCCGCGGACCTCATGCTGTCCGTTT-----GTGCACTTCG	2221
Oy	321	gMetLeuAlaSerGlnAlaSerAlaGIlyArgArgIeuSerAsp-----ValAr	337
Db	2220	CGAG-----CGCGGCGCGGTGAGGGGCGTGCATCCTCGACGCGAGGTGCG	2173
Oy	337	gGIlyGIlyIleuLeuArgLeuAlaGIlnAlaLeu-----AlaPheArgGlnAs	352





Db 28435 ATCTCTTCTGCTGCGCGCGCTGCTGCGAGCTGACATCCCACTACGCCGT 28494  
Qy 813 AAlaIleSerPheLeuGlnLysHisAlaAsp-----ProleuHis 825  
Db 28495 GCGAGCGCTTCTGCTGATGCGCTTGCACACCGCGCGCGCAAGGCGCTCCCTCA 28554  
Qy 826 AspleuSerPheAspAsn---SeraspLeuValMetLeuLysSerLeuLeuAlaGlyLeu 844  
Db 28555 TCGCTCGGCTGAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28614  
Qy 845 SerleuProSerArgAspAspArgThrAspArg---GlyLeuAspGlnGlnGlnGln 863  
Db 28615 GCGCGCGCT-----CGCATGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28662  
Qy 864 GluSerSerAlaGlySerleuProleuValSerValSerLeuPheThrProleuThr--- 882  
Db 28663 AGG-----CTCGCGCTGCTGATGGCGCGCTTCCGAACCGAGCGCGC 28707  
Qy 883 -----AlaAlaGlnMetAlaPro 888  
Db 28708 CTGCTCCCGCGCGCGCTGCTGACTTGAAGCGCGCTCAGCGCGCGCGCGCGCGCGCG 28767  
Qy 889 TyrMetLysArgLeuSerArgGlyGlnThrVal----- 899  
Db 28768 TTGTTCCACAGCTGCTGCTGCGCGCTGCGACCGCAAGCGCGCGCGCGCGCGCGCG 28827  
Qy 900 ---GluAspLeuLeuGlnValLeuSerAspIleAspGlnMetSerArgArgProGln 918  
Db 28828 GCGTCTGCTGCTGACAGCGCGCTGCTGACCGCTCCCGCGCGCGCGCGCGCGCGCGGT---GCC 28884  
Qy 919 IleleuSerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAla 934  
Db 28885 CTGCTGCATTCATTCACGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28932

## RESULT 5

US-08-456-837-6  
Sequence 6, Application US/08456837  
Patent No. 5643774

## GENERAL INFORMATION:

APPLICANT: Schnupp, Thomas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gaffney, Thomas Deane  
APPLICANT: Lam, Stephen Ting  
APPLICANT: Hammer, Phillip E.  
APPLICANT: Uknes, Scott Joseph  
TITLE OF INVENTION: Genes for the synthesis of  
NUMBER OF INVENTIONS: antipathogenic substances  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456.837  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: 08/457,205  
APPLICATION NUMBER: 01-JUN-1995  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8699  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-456-837-6

## Alignment Scores:

Pred. No.:	0.006	Length:	28958
Score:	150.50	Matches:	255
Percent Similarity:	34.668	Conservative:	111
Best Local Similarity:	24.158	Mismatches:	405
Query Match:	2.87%	Indels:	292
DB:	1	Gaps:	57

US-09-929-769-7 (1-1029) x US-08-456-837-6 (1-28958)

Qy 14 ThrLeuGlyProProArgAlaAspAspSerGluPheGlnAlaLeuLeuAsp----- 30  
Db 26222 ACCCTCGGCAACCGCTCTGCTGTC-TGAGCGCGACCGCGCGCTCTCGACGATGGCAT 26280  
Qy 31 IleTrpPheProGlnGlnLysProleuPro-----ThrAlaPheLeuValAspThrSer 48  
Db 26281 CGCTTCTTCTGCTGAGGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26340  
Qy 49 GlnGlnAlaLeuLeuLeuProAspTrpLeu---LysLeuArgMetLysArgSerGlnVal 67  
Db 26341 GAGCGCTACCGCTGATCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26400  
Qy 68 LeuArgLeuValAspAlaAlaLeuGlnAspLeuGlnProGlnGlnLeuLeuLeuVal 87  
Db 26401 GCGCGCTGCTGCTC---CTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26457  
Qy 88 GlnSerPheGlyIleProValSerMetSerLysLeuLeuGlnPheLeuAspGlnAla 107  
Db 26458 AACGCTTCTTCTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26508  
Qy 108 ValAlaHisAspProGlnThrLeuGlnGlnAsnIleMetAspLysAsnTrpMetAlaHis 127  
Db 26509 -----CCCTTCACAGCGCGCGCTTGTGCTGCGCGCGCGCGCGCGCGCGCGCG 26553  
Qy 128 LeuValGlnValGlnHisGlnArgGlyAlaSerGlyGlnThrPhe-----HisSer 145  
Db 26554 GCTGCGCGCGT-----GCCCTCGGAGCGCTGACCTGCGCGCGCGCGCGCGCGCG 26598  
Qy 146 LeuLeuThrAlaSerLeuProProArgArgAsp----- 156  
Db 26599 CTGCTCGGCG 26657  
Qy 157 -----SerThr 158  
Db 26658 CTCCCTCGAGAGACACCGCTGCTGAGACACCGCGCTTGGCATACCTGCTGCTGCG 26717  
Qy 159 GluAlaProLysProLysSerSerProGlu---GlnProIleGlyGlnGlyArgLeuArg 177  
Db 26718 CAGGCGCGCGCTCTGAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26777  
Qy 178 ValGlyThrGlnLeuArgValLeuGlyProGlnAspAspLeuAlaGlyMetPheLeuGln 197  
Db 26778 CGAAGACGTCAGCTCGA----- 26795  
Qy 198 IlePheProleuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeu 217

Db 26796 -----CCCCCCTCGCTCTCCATCGACGCGCGCTCTCTCCA--- 26837  
 QY 218 AlaLeuGlnAlaLeuGlnGluLeuAlaArgValValGlnGly---SerProGlu 236  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 26838 GATCTCGTGGGCGCGGACGAGTGTG-CTGGAC-----GAAGGCGCTCTCGCTTC 26887  
 QY 237 ValProGlyIleThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHis 256  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 26888 ATGACGCGCGCGCGACGAGCTCTTC-----AGATGCGCGCTCGACCTCGCCACAGCA--- 26938  
 QY 257 GlyValAlaLeuValMetSerMetHisArgSerHisPheLeuLacys----- 272  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 26939 ---GCGGCTCTCTCGCGCAAGCTAGCCGCTCC-CATTGCTTGATGCTCTCGCGAATGCC 26994  
 QY 273 ---ProLeuLeuArgLysLeuGlnGlyValArgCysValProGlnAspThr---Gly 290  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 26995 CCCCC-----TCGGGCGCGCACCGACGAGGACACCAAGT 27030  
 QY 291 PheSerSerLeuPheLeuLysValLeuLeuGlnMetLeuGlnTrpLeuAspSerProGly 310  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27031 TTCTAGCGACGC-----CTCGAGAGCGCTGCG 27057  
 QY 311 ValGlnGlyLysPro-----LeuArgAlaGlnLeuArgMet 322  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27058 CTTCGCTTATGGCCCGCATTCACAGGCGCTCGCGCGCTCTACAGCGCGCGAGCT 27117  
 QY 323 LeuAlaSerGlnAlaSerAlaGlyArgArg-----LeuSerAsp 335  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27118 CTTCGCGCGAAGCGCAGCTCCGCGCGCGCGCGCAAGAGCGCGCTTTTGGCCCTCCA 27177  
 QY 336 ValArgGlyLysLeuLeuArgLeuAlaGlnAlaLeuAlaPheArg----- 350  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27178 CCGCGCGCTCTCGACAGCGCTTCGAGCGCGCTCTTGTAGACAGCACAGCAAGGC 27237  
 QY 351 ---GlnAspLeuGlnValAlaSerSerThrValArgAlaValIleAlaThrLeuArgSer 369  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27238 CTTCAGATATCCCTTCTCTCGTGAAGCGAGT-----ATCGCTGGCGCTCGGCTTC 27285  
 QY 370 GlyGlnGlnCysSerValGluProAsp-----LeuIleSerLysValLeuGlnGly 386  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27286 -----GAGCCACACCCCTGGCGCGCTTTCACACGCTCTGAGGCG 27327  
 QY 387 LeuIleGlnValArgSerProHisLeuGlnGluLeuLeuThrAlaPhePheSerAlaThr 406  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27328 AATCTCGCGCTCGCTCC-----TCTCTCCCG 27354  
 QY 407 AlaAspAlaAlaSerProPheProAlaCysLys-----ProValValAlaValSerSer 424  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27355 ACGCCAGAGCGCAACCATCGCTCGGTGCAGCGCTCGCATCGCG----- 27402  
 QY 425 LeuLeuLeuGlnGlnGluProLeuAlaGlyLysProGlyAlaAspGlyLys 444  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27403 -----CCGCGTCCGCGAGCGAGCTCGCGAGACCGCGGAGCG-----TCCAC 27444  
 QY 445 LeuGlnAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuGlnMetLeu 464  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27445 CTGATATCCCTC-----TTCGATCATCAGCTGAGACGACTCA- 27482  
 QY 465 AspProGluValAlaSer-----SerCysProAspLeuGlnLeuArgLeuLeu 480  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27483 AACCCCACTACACCGCGCATCGCGCGCGAGCGGTGCTCTCGCGCACAGAGS----- 27536  
 QY 481 PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProGlyLeuLeu 500  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27537 ---TTCGACCTCGGAGCGAGGT---GCCTTCGACCGCTATACGACCTGTGTCT 27590  
 QY 501 Thr-----LeuPheThrHisGlnSerSerTrpTrpThrLeuHisGln 514  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27591 ACSCAGCGCCCTCGACAGGCGCTTCCTCCCAAGCCTCGATCGCCCTTCATACGC 27650  
 QY 515 CysAlaLeuValLeuLeuGlnGlyLysSerArgGln-----ArgPheAspPro 530  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27651 TCTGCCGGAAGGAGACTCTATCTCGAGAGCGCGCGAGACCAACCGCGCTCGCT 27710

QY 531 SerAlaSerLeuAspPheLeuTrpAlaCysIleHisValProArgIleTrpGlnGlyArg 550  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27711 CTTCGAAGCGCTG-----GCTCGCGAGAGAGCGCTCGCTCTCGCGCGCTCG 27758  
 QY 551 AspGlnArgThrProGlnLysArgArgGlnGluLeuValLeuArgValGlnGlyProGlu 570  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27759 ---CCTGCTACCGAGCGCGCGCTCGCGCACCGCTGAGAGAGCTCAAGGCGCTCG- 27814  
 QY 571 LeuIleSerLeuValGlnLeuIleLeuAlaGlnAlaGlnThrArgSerGlnAspGlyAsp 590  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27815 CTCAGAGCGCTCTCTGGGGCTCGCTCGCTCGCGCGAGAGAGACCAAGAGCGGCTTC 27874  
 QY 591 ThrAlaAlaCysSerLeuIleGlnAlaArgLeuProLeuLeuLeuSerCysCysGly 610  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27875 TCGTCTCTGACCTCGACAGCGAGCGAGCGCTCCAGCAGCGCTCTCGCGCGCTCG 27934  
 QY 611 AspAspGlnSerValArgLysValThrGlnHisLeuSerGlyLysIleGlnGlnTrpGly 630  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27935 ACSCAGAGAGCGAGAGATCGCTCGCAAGCGCAACCCCTCGTTCAA-----GGC 27988  
 QY 631 AspSerValLeuGlyArgArgCysArgAspLeuLeuGlnLeuTrpLeuGlnArgPro 650  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 27989 TCTCAGCGCTCGCGCGAGCGCGCGCG-----ACACAGCGCTCC 28027  
 QY 651 GlnLeuArgValProValProGlnValLeuLeuHisSerGlnGlyAlaAlaSerSer 670  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28028 CCGAGAGCTCTCGAGAGCACCGTCTCCTACAGGAGGACCGCGAGCTCGCGCGCTCG 28086  
 QY 671 ValCysLysLeuAspGlyLeuIleHisArgPheIleThrLeuLeu----- 685  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28087 GTGCGCGCGCGCTCTGTGTAACACAGAGCGCAAGCACCTGCTCTCACTCGCGCGAG 28146  
 QY 686 -----AlaAspThrSerAspSerArgAlaLeuGlnAlaAspArgGlyAla 699  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28147 GCGCGAGCGCTCGCGGTGCTGATGCTCTCGGAAGCAG-----CTCGAAGCTTGGGGCT 28203  
 QY 700 AspAlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeu---LeuLeuArg 718  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28204 TCGATTCACCTCGCGCGCTCGAGCGAGTGCAGCGCGCTCAAGGAGCACTTGTGAT 28263  
 QY 719 HisLeuProMet-----IleAlaAlaLeuLeuHisGlyArgThrHisLeuAspPhe 735  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28264 AACATTCGAGCGCTCACCGCGGTGCGCGCTGTGATCGCGCGAGCTCTGAGAGC 28323  
 QY 736 GlnGluPheArgGlnGlnAlaHisLeuSerCysPheLeuHisValLeuGlnGlyLeuGlu 755  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28324 GAT-----CTGCTCGCGCGCGCATGAGC 28344  
 QY 756 LeuLeuGln---ProHisValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeu 774  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28345 CTCGAGCGGATCGACCGCGCTTCGCGCGCAAGATCGATCGCGC---TGGCAC----- 28395  
 QY 775 LeuSerPheIleArgLeuLeuLeuAsnTrpArgLysSerSerArgHisAlaAlaPhe 794  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28396 -----TTCATCACTACACCAATAGAGCCCTTCCGCGCTTC 28434  
 QY 795 Ile-----AsnLysPheValGlnPheIleHisLysTrpIleThrTrpAsnAlaProAla 812  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28435 ATCCTCTTCGTCGTCGCGCGCGCTCTCGCGAGTCAAGTCACTCACTACAGCGCT 28494  
 QY 813 AlaIleSerPheLeuGlnLysHisAlaAsp-----ProLeuHis 825  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28495 GCGAGCGCTCTCTCGATGCGCTTGGCGCACACCGCGCGCGAGGCGCTCTGCTCA 28554  
 QY 826 AspLeuSerPheAspAsn---SerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeu 844  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28555 TCGTCTGCTGAGAGCACTGCGCGAGCGAGCGCAATGACAGACAGCTCAAGCGCGCC 28614  
 QY 845 SerLeuProSerArgAspAspArgThrAspArg---GlyLeuAspGlnGlnGlyGln 863  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 28615 GCGCGCGCT-----CCATGAGAGCGCGCGCTTCCCTGACCTCTGAGAGAG 28662

```

QY 864 GluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThr--- 882
Db 28663 AGG-----CTGGCCCTCTTCGATGCGGGCTCTTCGACAGCGAGCCGCC 28707
QY 883 -----AAlaAlaGluMetAlaPro 888
Db 28708 CTGGTCCCGCGGCTTCGACTTGAGCGCGCTCAGGCGGAGACGCCGCGAGCTCCCGC 28767
QY 889 TyrMetLysArgLeuSerArgGlyGlnThrVal----- 899
Db 28768 TTGTTCCAGCTCTGCTCCGCGCTCCGACCGTACCGACGAGCCGCCAGACACCGCCAG 28827
QY 900 ---GluAspLeuLeuGluValLeuSerAspIleAspGluMetSerArgArgArgProGlu 918
Db 28828 GCGTCTGCGCTTACAGAGCGGCTCTCAGCCCTCCGCCGCCGAGAGCGGAGCGT---GCC 28884
QY 919 IleLeuSerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAla 934
Db 28885 CTGCTCGATCTCATCCGACCGAGCCGCCGCTCTCCGCGCTCGCC 28932

```

## RESULT 6

```

US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO

```

```

; ANTI-SENSE: NO
US-08-457-342-6
Alignment Scores:
Pred. No.: 0.006 Length: 28958
Score: 150.50 Matches: 255
Percent Similarity: 34.66% Conservative: 111
Best Local Similarity: 24.15% Mismatches: 405
Query Match: 2.87% Indels: 292
DB: 1 Gaps: 57

```

US-09-929-769-7 (1-1029) x US-08-457-342-6 (1-28958)

```

QY 14 ThrLeuGlyProProArgAlaAspSerGluPheGlnAlaLeuAsp----- 30
Db 26222 ACCTCCGGCAACCGCTCTGTTTC-TGAGCGCGAGACCGAGCGGCTCTTGACGATGGGAT 26280
QY 31 IleTrpPheProGluGluLysProLeuPro-----ThrAlaPheValAspThrSer 48
Db 26281 CGTTCTTCGTCGAGGTCAGCGCCCATCCGTCGTCAGCTCGGCTCCGAGACCTGC 26340
QY 49 GluLysAlaLeuLeuProAspThrLeu---LysLeuArgMetIleArgSerGluVal 67
Db 26341 GAGGCTCACCGCTCGATCCGTCGCTCGCTCCATTCGACGCGAGAGCCACCTC 26400
QY 68 LeuArgLeuValAspAlaAlaLeuGlnAspLeuGluProGlnGlnLeuLeuPheVal 87
Db 26401 GCCCGCGCTC-----CTCTCCGCGGAGAGCTCTACCCAGCGCTCGCCCTGCAGCTG 26457
QY 88 GlnSerPheGlyIleProValSerMetSerLysLeuGlnPheLeuAspGlnAla 107
Db 26458 AACCCCTTCCTGCGCGCTCCGCCGCAAGCTCTCCACCTAC----- 26508
QY 108 ValAlaHisAspProGlnThrLeuGlnGlnAsnIleMetAspLysAsnTyrMetAlaHis 127
Db 26509 -----CCCTTCAACGAGGCGCTTGGCTGAGCGCTC---ACGCGCAC 26553
QY 128 LeuValGluValGlnHisGluArgIleValAspArgGlyGlnThrPhe-----HisSer 145
Db 26554 GCTGCGGAGCTC-----GCTCCGAGGCTGACCTCGCGGACGACCGG 26598
QY 146 LeuLeuThrAlaSerLeuProProArgArgAsp----- 156
Db 26599 CTGCTCGCGCGCGCGCTGCG---CTCGCGGACCGGAGCTTGTCTTCACAGAGCGCT 26657
QY 157 -----SerThr 158
Db 26658 CTCCCTCGCAGAGCACCCGCTGCTCGAAGACACCTGCTTCGACATACCTGTCTGC 26717
QY 159 GluAlaProLysProLysSerSerProGlu---GlnProIleGlyGlnIleArgIleArg 177
Db 26718 CAGGCGCGCGCTCGCGAGCTCGGCTGCATGTCGCCATCGTCGCGCTCGACACGCT 26777
QY 178 ValGlyThrGlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGln 197
Db 26778 CGAAGACGTACAGCTCGA----- 26795
QY 198 IlePheProLeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeu 217
Db 26796 -----CCCCCCTGCTCTCTCCATCGAGGCGCGCTCTCTCCA--- 26837
QY 218 AlaLeuGlnAlaLeuGlyGlnIleLeuAlaArgValValGlnGly---SerProGlu 236
Db 26838 GATCCGCTCGGCGCGCGGAGCGGTCGAGAC-----GAAAGCGGCTCTCCCTC 26887
QY 237 ValProGlyIleThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHis 256
Db 26888 ATAGCGCGCGCGACAGCGCTTC-----AGATAGCCCTCGAGCTCGACGCCA--- 26938
QY 257 GlyGlyAlaLeuValMetSerMetHisArgSerHisPheLeuAlaCys----- 272
Db 26939 ---GGGCTCTCTCGCGGACGACCTGACCGCTCC-CATTGCTTCGATGCTCCGCAATGCG 26994

```



QY	273	---	ProLeuLeuArgGlnLeu	CysGlnTrpGlnArgCysVal	ProGlnAspThr---	Gly	290
Db	26995	CCCCC-----	-----	-----	-----	-----	-----
QY	291	PheSerSerLeuPheLeu	ValLeuLeuGlnMetLeu	GlnTrpLeuAspSerProGly			310
Db	27031	TTCTACGCAGCC-	-----	-----	-----	CTCAGAGCCGCTGG	27037
QY	311	ValGlnGlyGlyPro-	-----	-----	-----	LeuArgAlaGlnLeuArgMet	322
Db	27058	CTTCCTTAATGGCCCGAGTTC	CCAGGGCCCTCCGCCCGGCTCTACAGCGGGCGAGAGCT				27117
QY	323	LeuAlaSerGlnAlaSerAla	GlyArgArg-----	-----	-----	LeuSerAsp	335
Db	27118	CTTCGCCGAAGCCAAAGCTCCCGAGCCCGCGAAGAGACGCCGCTGCTTTTGCCCTCA					27177
QY	336	ValArgGlyGlyLeuLeu	ArgLeuAlaGlnAlaLeuAlaPheArg-----	-----	-----		350
Db	27178	CCCCCGCTGCCTCGACAGCGCTTCAGAGGCTCCGCTTTGTAAGACAGCAAGCAAAGC					27237
QY	351	---GlnAspLeuGlnVal	ValSerSerThrValAlaArgAlaValAlaLeuAlaThrLeuArgSer				369
Db	27238	CTTCAGAGATGCCCTCTCGTGAGCGAGAT-----	ATCGTGGCGCTCCGGTGC				27285
QY	370	GlyGlnGlnCysSerValGlu	ProAsp-----	-----	-----	LeuIleSerLysValLeuGlnGly	386
Db	27286	-----	-----	-----	-----	-----	-----
QY	387	LeuIleGlnValArgSerPro	IleuGlnGlnLeuLeuThrAlaPheSerAlaThr				406
Db	27328	AATCTCCGCGCTCC-----	-----	-----	-----	-----	27354
QY	407	AlaAspAlaAlaSerPro	PheProAlaAlaCysLys-----	-----	-----	ProValValValSerSer	424
Db	27355	ACGCCAGAGCGAACCCTACGCTCGGTGCAAGCGCTCGCATGGCG-----	-----	-----	-----	-----	27402
QY	425	LeuLeuLeuGlnGlnGln	ProLeuAlaGlyLysProGlyAlaAspGlyLys				444
Db	27403	-----	-----	-----	-----	-----	27444
QY	445	LeuGlnAlaValArgLeu	GlyProSerSerGlyLeuLeuValAspTrpLeuGlnMetLeu				464
Db	27445	CTCGATGGCTC-----	-----	-----	-----	-----	27482
QY	465	AspProGluValValSer-	-----	-----	-----	SerCysProAspLeuGlnLeuArgLeu	480
Db	27483	AAGCCCAACCTACCGCCCATCGCCCGACGCGTCCCTCCGCGACAGAGG-----	-----	-----	-----	-----	27536
QY	481	PheSerArgArgGlyGlyLys	GlyAlaGlnValProSerPheArgProIyrLeu				500
Db	27537	---TCTGCACTCGGGACAGAGGT---	-GCCTCGACCGCTATACCGCACTTGCTGCTCT				27590
QY	501	Thr-----	-----	-----	-----	LeuPheThrHisGlnSerSerTrpProThrLeuHisGln	514
Db	27591	ACGCAGGCCCTCGACCAAGGGCGCTTCGCCCTCCCAAGCTCGTCAATCGCCCTTATCGC					27630
QY	515	CysIleArgValLeuLeu	GlyLysSerArgGln-----	-----	-----	ArgPheAspPro	530
Db	27651	TCTGCCCGAAGCGACCTCATCGCAGCGCCGCGAGACACCGCGACAGCGCTCGCCCT					27710
QY	531	SerAlaSerLeuAspPhe	LeuTrpAlaCysIleHisValProArgIleTrpGlnIyrArg				550
Db	27711	CTTGACACCTG-----	-----	-----	-----	-----	27758
QY	551	AspGlnArgTrpPro	GlnLysArgArgGlnGlnValLeuArgValGlnGlyProGln				570
Db	27759	---CCTCGTCAACCGACGCGCGCTGCGCACACCGCTCAAGAGACGTCAAGGGCTCG-					27814
QY	571	LeuIleSerLeuValGluLeu	IleLeuAlaGlnAlaGlnIleThrArgSerGlnAlaAspLys				590
Db	27815	CTCACGCGCCCTCTGGGGGTCTCGGTGCTCGCGGCAAGGAGACCAAGCGCCCTC					27874
QY	591	ThrAlaAlaCysSerLeu	IleGlnAlaArgLeuProLeuLeuLeuSerCysCysCysGly				610

Db	27875	TCGTCCTCTCGACGCGACGAGGAGGAGGAGCCCTCCAGCAGCCCTGCTCGGCGCGCTCG	27934
QY	611	AspAspGluSerValArgLysValThrGluHisLeuSerGlyCylLeuIleGlnIleProGly	630
Db	27935	ACGCAAGAACGACCAAGATCGCCCTCCGACAGCAACCCCTGCTTCCA-----GGC	27988
QY	631	AspSerValLeuGlyArgGlyCysArgAspLeuLeuGlnLeuTyrLeuGlnIleArgPro	650
Db	27989	TCCTCAAGCCTGCCCGACGAGCGCCACAG-----ACACAGCGTCCC	28027
QY	651	GluLeuArgValProValProGluValLeuLeuHisSerGlnGlyAlaIleSerSer	670
Db	28028	CCGCAAGGCTCCGAGGACCGCTC-CTCATCAGGAGGAGGACCGGACGCTCGGCGCTCG	28086
QY	671	ValCysLysLeuAspGlyLeuLeuLeuHisArgPheIleThrLeuLeu-----	685
Db	28087	GTTCGGCGCGCGCTGCTGCTGAACCAACAGCCGCAACGACCTGCTCTCACTCGGCGCG	28146
QY	686	-----AlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAla	699
Db	28147	GCGCGAGGCGCTCCGGGCTGCTGATGCTCTGCAAGCGAG---CTCGAAGCTCTGGGGCT	28203
QY	700	AspAlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeu---LeuLeuArg	718
Db	28204	TCGGTCACCTCTGCGCGCTGCGACGTGCGGATCCAGCGGCTGTAAAGGACTCTGTGAT	28263
QY	719	HisLeuProMet-----IleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPhe	735
Db	28264	AACATTTCCGAGCGCTCACCCCGGTCGCGCGCTCGTGAGATCGCGGACGCTCTCGACGGC	28322
QY	736	GlnGluPheArgGlnGlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGln	755
Db	28324	GAT-----CTCTCGCGCCCATGAGC	28344
QY	756	LeuLeuGln---ProHisValPheArgSerGlnHisGlnGlyAlaLeuArgPaspCysLeu	774
Db	28345	CTCGACGCGATGACCGCGCTCTTCGCCCCCAAGATCATGATCGCGCC---TGGCAC-----	28395
QY	775	LeuSerPheIleArgLeuLeuAsnTyrArgLysSerSerArgHisLeuAlaPhe	794
Db	28396	-----TTCGATCAGCTCACCCCAAGATAGCCCTTCCGCGCTTC	28434
QY	795	Ile-----AsnLysPheValIlePheIleHisLysTyrIleThrTyrAsnIleProIle	812
Db	28435	ATCTCTTCTCTCGCTGCGCGCGCTCTCTCGACAGCTCAAGTCACTCCAACTACGCGCT	28494
QY	813	AlaIleSerPheLeuGlnLysHisAlaAsp-----ProLeuHis	825
Db	28495	GCGACGCGCTCTCGATGCGCTTCCGCGCACCGCGCGCGCGCAAGGCGCTCGCTGCTCA	28554
QY	826	AspLeuSerPheAspAsn---SerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeu	844
Db	28555	TCGCTCGGTGGAGCACCTCGGGCCAGCGACGCAATGACAGACACGACCTAGCGCGCCG	28614
QY	845	SerLeuProSerArgAspAspArgThrAspArg---GlyLeuAspGlnGluGlyGlnGln	863
Db	28615	GCGCGCCCT-----CGAATGGAGCGCGCGCGCTTCCCTCGACGCTTGAGGAG	28662
QY	864	GluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThr---	882
Db	28663	AGG-----CTCGCCCTCTTCGATGCGCGGCTTTCGCAACGACGACGCGC	28707
QY	883	-----AlaAlaGluMetIleLapro	888
Db	28708	CTGGTCCCGCGCGCTTGACTTGAGCGCGCTCAGGCGAAGCCCGGCGAGCGTCCCGCG	28767
QY	889	TyrMetLysArgLeuSerArgGlyGlnThrVal-----	899
Db	28768	TTGTCTCAACGCTCTGCTGCGGCTTGCAACCGGCAAGCGCGCGCGCAACACGCGCCAG	28827
QY	900	---GluAspLeuLeuGlnValLeuSerAspLysGlnMetSerArgArgArgProGlu	918



```

Db 28828 GCCTGCTGCTTACAGAGGGCCTCTCAGGCCCTCCGGCCCGCCGAGCAGGAGCGT---GCC 28885
Oy 919 lIleuSerPhepheSerThrAsnIeuGlnArgIeuMetSerSeraIa 934
      :::::
Db 28885 CTGCTCGATCTCATTCACGACCGAGCCGCCGCTGCTCGCCCTCGCC 28932
      :::::

RESULT 7
US-08-457-646A-6
: Sequence 6, Application US/08457646A
: Patent No. 5679560
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligott, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Clba-geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,646A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEO ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-457-646A-6

Alignment Scores:
Pred. No.: 0.006 Length: 28958
Score: 150.50 Matches: 255
Percent Similarity: 34.66% Conservative: 111
Best Local Similarity: 24.15% Mismatches: 405
Query Match: 2.87% Indels: 292
DB: 1 Gaps: 57

US-09-929-769-7 (1-1029) x US-08-457-646A-6 (1-28958)
Oy 14 ThrIeuGlyProIleArgIAlaSpaSpSerGluPheGlnAlaIleuLeuAsp-----30

```

Db 27118 CTTCCGCGAGCGCAAGCTCCCGAGCGCGCGAGAGAGCGCGCTGCTTTGCCCTCCA 27177  
 Oy 336 ValArgGlyGlyLeuLeuArgLeuAlaLeuAlaPheArg----- 350  
 Db 27178 CCCCCTGCTGCGACAGCGCTTTCAGGCGCTCGCTTTGAGACGACGCAAGGCG 27237  
 Oy 351 ---GlnAspLeuGluValValSerSerThrValArgAlaValIleAlaThrLeuArgSer 369  
 Db 27238 CTTACAGATGCTCTGCTGTGAGCGAGT-----ATGCGTGGCTCCGCTCG 27285  
 Oy 370 GlyGluGlnGlySerValGluProAsp-----LeuIleSerTyrValLeuGlnGly 386  
 Db 27286 -----GAGCCACCGACCGCTGCGCTGCTTCCACCGCTGTGAGGCG 27327  
 Oy 387 LeuIleGluValArgSerProHisLeuGluGluLeuThrAlaPhePheSerAlaThr 406  
 Db 27328 AATCTCGGCTGCTC-----TCCTCGCGG 27354  
 Oy 407 AlaAspAlaIleSerProPheProAlaCysLys-----ProValValValSerSer 424  
 Db 27355 ACGCAGAGCGCAACCATCGCTCGTCAAGCGCTCGCATGCGCG----- 27402  
 Oy 425 LeuLeuLeuGlnGluGluProLeuAlaGlyGlyLysProGlyAlaAspGlySer 444  
 Db 27403 -----CCGCGTCCGCGAGCACCTCCGACACCGCGGAGCG-----TCCAC 27444  
 Oy 445 LeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspThrLeuGluMetLeu 464  
 Db 27445 CTCGATGCGCTC-----TTCCGATGCGATCGAGCGAGCGTGCAT- 27482  
 Oy 465 AspProGluValValSer-----SerCysProAspLeuGlnLeuArgLeuLeu 480  
 Db 27483 AAGCCCATCTACCGCATCGCCCGAGCGGCTGCTCTCCGACAGAGG----- 27536  
 Oy 481 PheSerArgArgGlyGlyGlnAlaGlnValProSerPheArgProTyrLeuLeu 500  
 Db 27537 ---TCTCGACCTCGGACCAAGGT---GCTCTCGACCGCTATACGACCTGCTCTCT 27590  
 Oy 501 Thr-----LeuPheThrHisGlnSerSerThrProThrLeuHisGln 514  
 Db 27591 ACGCAGCGCTCGACACGAGCGCTCGCTCCAAAGCTCGTATGCGCCCTTCATCGC 27650  
 Oy 515 CysIleArgValLeuLeuGlyLysSerArgGluGln-----ArgPheAspPro 550  
 Db 27651 TCTGCCGAGGAGCATCATCGAGCGCGCGAGACCGCGACGCGCTGCGCT 27710  
 Oy 531 SerAlaSerLeuAspPheLeuThrAlaCysIleHisValProArgIleTyrGlnArg 550  
 Db 27711 CTTGCAAGCTG-----GCTGCCGAGAGCGCGCTGCTCGCGCGCTCGC 27758  
 Oy 551 AspGlnArgThrProGlnLysArgArgGluLeuValLeuArgValGlnGlyProGlu 570  
 Db 27759 ---CCTCGTACCGCGAGCGGCTCGCGCACCCAGCGTGAAGAAGAGTCAAGGCGCTCG- 27814  
 Oy 571 LeuIleSerLeuValGluLeuIleLeuAlaGluAlaGluThrArgSerGlnAspArg 590  
 Db 27815 CTCACGCGCTCTGCGGATCTCGCTCGCTCGCGGAGAGGAGCACCGACAGCGCTC 27874  
 Oy 591 ThrAlaAlaCysSerLeuLeuGlnAlaArgLeuProLeuLeuSerCysCysGly 610  
 Db 27875 TGTCTCTCTGACCTCGACGAGAGGAGCGCTCCAGACGCGCTCGCGCGCTCG 27934  
 Oy 611 AspAspGluSerValArgLysValThrGlnHisLeuSerGlyCysIleGlnGlnTyrGly 650  
 Db 27935 ACGCAAGAGCGAGATCGCTCGCGACGCGCAAAACCCCTCGTTCAA-----GCC 27988  
 Oy 631 AspSerValLeuGluArgArgCysArgAspLeuLeuGlnLeuTyrLeuGlnArgPro 650  
 Db 27989 TCTCAGCTGCGCCAGGCGCGCACG-----ACACGCGCTCGC 28027  
 Oy 651 GluLeuArgValProValProGluValLeuLeuHisSerGlyGlyAlaAlaSerSer 670  
 Db 28028 CCGCAGGCGCTCGAGGCGACCGTCTCATATCAGGAGGAGGACCGCGACGCTCGCGCGCTG 28086

Oy 671 ValCysLysLeuAspGlyLeuLeuIleIleArgPheIleThrLeuLeu----- 685  
 Db 28087 GTCCGCGCGCGCTGCTGTAAACGACGAGCGCAACGACCTGCTGCTCACTCGCGCG 28146  
 Oy 686 -----AlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAla 699  
 Db 28147 GCGCGAGCGCTCCGGGCTGTATGTTCTGCAAGGAG---CTCGAACCTGTGGGCGCT 28203  
 Oy 700 AspAlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeu---LeuLeuArg 718  
 Db 28204 TCGGTACCTGCGCGCGCTGCGACGCGGATGCGGATGCCAGCGCTTAAGAGACTTGTGAT 28263  
 Oy 719 HisLeuProMet-----IleAlaAlaLeuHisGlyArgThrHisLeuAsnPhe 735  
 Db 28264 AACATTCCAGCGCTCACCGCGCTGCGCGCGCTGCTCATGCGCGCGAGCTGCTCGAGCG 28323  
 Oy 736 GlnGluPheArgGlnGlnHisHisLeuSerCysPheLeuHisValLeuGlyLeuGly 755  
 Db 28324 GAT-----CTGCTCGCGCGCATGAGC 28344  
 Oy 756 LeuLeuGln---ProHisValPheArgSerGlnHisGlnGlyAlaLeuThrAspCysLeu 774  
 Db 28345 CTCGACGAGATGACCGCGCTTTCGCGCGCGCGCGAGATGATGCGCG---TGGCAC 28395  
 Oy 775 LeuSerPheIleArgLeuLeuLeuAsnTyrArgLysSerSerArgHisLeuAlaPhe 794  
 Db 28396 -----TTCATCAGCTCACCGACCAAGATMAAGCCCTTGGCGCGCTTC 28434  
 Oy 795 Ile-----AsnLysPheValGlnPheIleHisLysTyrIleThrTyrAsnAlaProAla 812  
 Db 28435 ATCTCTCTGCGCTGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCT 28494  
 Oy 813 AlaIleSerPheLeuGlnLysHisAlaAsp-----ProLeuHis 825  
 Db 28495 GCGAGCGCGCTGCTGATGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCT 28554  
 Oy 826 AspLeuSerPheAspAsn---SerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeu 844  
 Db 28555 TCCCTGCGCTGCG 28614  
 Oy 845 SerLeuProSerArgAspArgThrAspArg---GlyLeuAspGluGlnGlyGlnGlu 863  
 Db 28615 GCGCGCGCT-----GCGATGAGAGCGCGCGCGCTTCCGCGCGCTGAGAG 28662  
 Oy 864 GluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThr 882  
 Db 28663 AGG-----CTCGCGCTTTCGATGCGCGCGCTTTCGGAACCGAGACCGCG 28707  
 Oy 883 -----AlaAlaGluMetAlaPro 888  
 Db 28708 CTGCTCCCGCGCGCTTCGACTTGAAGCGCGCTGAGGCGCAACCGCGACGCTCCCGCG 28767  
 Oy 889 TyrMetLysArgLeuSerArgGlyGlnThrVal----- 899  
 Db 28768 TTGTTCCAAAGCTTCTGCTGCGCGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCG 28827  
 Oy 900 ---GlnAspLeuGluValLeuLeuSerAspIleAspGluMetSerArgArgArgProGlu 918  
 Db 28828 GCGTCTGCTTACAGAGCGCGCTTCAGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCT 28884  
 Oy 919 IleLeuSerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAla 934  
 Db 28885 CTGCTGATCTCATCGCACCGAGCGCGCGCGCTGCTGCGCGCTGCGCGCTGCGCG 28932  
 RESULT 8  
 US-08-458-076A-6  
 ; Sequence 6, Application US/08458076A  
 ; Patent No. 5698425  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James M.  
 ; APPLICANT: Beck, James Joseph





COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA: US/08/764,233A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/729,214  
 FILING DATE: 09-OCT-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/258,261  
 FILING DATE: 08-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: 1506/CIP6  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919) 541-8587  
 TELEFAX: (919) 541-8689  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Sorangium cellulosum  
 IMMEDIATE SOURCE:  
 CLONE: p98/1  
 US-08-764-233A-4

Alignment Scores:  
 Pred. No.: 0.006 Length: 28958  
 Score: 150.50 Matches: 255  
 Percent Similarity: 34.66% Conservative: 111  
 Best Local Similarity: 24.15% Mismatches: 405  
 Query Match: 2.87% Indels: 292  
 DB: 1 Gaps: 57

US-09-929-769-7 (1-1029) x US-08-764-233A-4 (1-28958)  
 QY 14 ThrleuglProProArgAlaAspSerGluPheGlnAlaLeuAsp----- 30  
 DB 26222 ACCCTCGGGAACCGTCGCTTC-TGAGCGGAGACGCGCTCTCGACATGGGCA 26280  
 QY 31 lletPheProGluGluysProLeuPro-----ThrAlaPheLeuValAspThrSer 48  
 DB 26281 CGCTTCCTGTCGAGGTGACGCGCCATCCCTGCTCAGCGCTCGCCCTCGGAGACCTGC 26340  
 QY 49 GluGluAlaLeuLeuProAspThrLeu---LysLeuArgMetLeuArgSerGluVal 67  
 DB 26341 GAGCGCTCAGCGCTGATGCCCTGCTGCTGCTCCATTCGAGCGGAGACGACGACCTTC 26400  
 QY 68 LeuArgLeuValAspAlaLeuGlnAspLeuGluProGlnGlnLeuLeuPheVal 87  
 DB 26401 GCCCGCTGCTC---CTGCTCGGGGAGGAGCTCTACCGGAGGCTCGCGCTCGACCTGC 26457  
 QY 88 GlnSerPheGlyIleProValSerSerMetSerLysLeuLeuGlnPheLeuAspGlnAla 107  
 DB 26458 AACGCTTCCTCGCGCTCGCTCGCTCGCGCAAGTCTCCCTCCACCTAC----- 26508  
 QY 108 ValAlaHisAspProGlnThrLeuGlnGlnAsnIleMetAspLysAsnThrMetAlaHis 127  
 DB 26509 -----CCCTTCACACGCGACGCGCTTCCTGCTCGAGCGCTCC---ACGGCGCAC 26553  
 QY 128 LeuValGluValGlnHisGlnArgLysAlaSerGlyGlnThrPhe-----HisSer 145  
 DB 26554 GCTCGGACGCTC-----GCTTCGCGAGGCTGACCTCGGCGCACACCGCG 26598  
 QY 146 LeuLeuThrAlaSerLeuProProArgArgAsp----- 156

DB 26599 CTGCTCGGGCGCGCGCTGC-CCTGCGGACGCGATGAGCTTTGCTTCACAGACGCGT 26657  
 QY 157 -----SerThr 158  
 DB 26658 CTCCTCGGAGACACCGCTGGCTCGAGACAGCTGCTTCGGCATACCTGCTCTCC 26717  
 QY 159 GluAlaProLysProLysSerSerProGlu---GlnProIleGlyGlnArgLeuArg 177  
 DB 26718 CAGCGCGCGCTCTCGAGCTGCGCTCGATGCGCCATCTGCTGCGCTCGACACCT 26777  
 QY 178 ValGlyThrGlnLeuArgValLeuGlyProGluAspLeuAlaGlyMetPheLeuGln 197  
 DB 26778 CCAAGACGTCAGCTCGA----- 26795  
 QY 198 IlePheProLeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeu 217  
 DB 26796 -----CCCCCTCGCTCTCCATCGACAGGCGCGCTCTCTCA--- 26837  
 QY 218 AlaLeuGlnGlnAlaLeuGlyGlnLeuAlaArgValValGlnGly---SerProGlu 236  
 DB 26838 GATCTCGTGGGCGCGCGGACGCTG-CGTGAC-----GAAGCGGCTCTCGCTTC 26887  
 QY 237 ValProGlyIleThrValArgValLeuGlnAlaLeuAlaThrLeuSerSerProHis 256  
 DB 26888 ATAGCGGCGCGCACGCGGCTTC-----AGATGCGCGCTGACTCCGACGCA--- 26938  
 QY 257 GlyGlyAlaLeuValMetSerMetHisArgSerHisPheLeuAlaGly----- 272  
 DB 26939 ---GGGCTCTCTCGCGACGCTAGCCGCTCC-CATTGCTCTCGATGCTCCGGAATGCG 26994  
 QY 273 ---ProLeuLeuArgGlnLeuGlyGlnTrpGlnArgCysValProGlnAspThr---Gly 290  
 DB 26995 CCCCC-----TCGGGCGCACCGACGCTGACACCAAGCT 27030  
 QY 291 PheSerSerLeuPheLeuLysValLeuLeuGlnMetLeuGlnTrpLeuAspSerProGly 310  
 DB 27031 TTCTACGCGCGC-----CTCGAGACGCTGG 27057  
 QY 311 ValGluGlyGlyPro-----LeuArgIleGlnLeuArgMet 322  
 DB 27058 CTTCGCTTATGCGCGCGAGTCCAGGCGCTCCGCGCGCTCTTCACAGCGGCGACGAGCT 27117  
 QY 323 LeuAlaSerGlnAlaSerAlaGlyArgArg-----LeuSerAsp 335  
 DB 27118 CTTCGCGGAAGCAACCTCCCGAGCGCGCGCAAGAGACGCGCGCTTTGGCTCCCA 27177  
 QY 336 ValArgGlyGlyLeuLeuArgLeuAlaGlnAlaLeuAlaPheArg----- 350  
 DB 27178 CCGCGCGCTGTCGACAGCGCTTGCAGCGCGCTTGTAGAGACGACGACCAAGCG 27237  
 QY 351 ---GlnAspLeuGlnValValSerSerThrValArgAlaValIleAlaThrLeuArgSer 369  
 DB 27238 CTTCAGAGATCCCTTCTCTGAGCGGAGT-----ATCGCTCGCTCGCGTGC 27285  
 QY 370 GlyGluGlnCysSerValGluProAsp-----LeuIleSerLysValLeuGlnCyl 386  
 DB 27286 -----GAGCGACACCGCTGCGGCTTCCACCGTCTCGAGGCGG 27327  
 QY 387 LeuIleGluValArgSerProHisLeuGlnGluLeuLeuThrAlaPheSerAlaThr 406  
 DB 27328 AATCTCGCGCTCGCTCC-----TCCTCGCGG 27354  
 QY 407 AlaAspAlaAlaSerProPheProAlaCysLys-----ProValValValAlaSerSer 424  
 DB 27355 ACCCGAGAGCGGACCAATCGCTCGGTCCAGACGCTCGCCATGCGCG----- 27402  
 QY 425 LeuLeuLeuGlnGluGluProLeuAlaGlyLysProGlyAlaAlaSerLysLys 444  
 DB 27403 -----CCGCGTCCCGCGAGACCTCGGACGCGGAGCG-----TCGCCAC 27444  
 QY 445 LeuGluAlaValAlaArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuGluMetLeu 464  
 DB 27445 CTGATGCGCTC-----TTCCGATCGACTGAGCGAGCTGCA- 27482

QY	465	AspProGluValValSer-----SerCysProAspLeuGlnLeuLeuLeuLeu	480
Db	27483	AAAGCCCACTACCGGCGCATGGCCCCCGAGCGCGCCCTCCGCGACAGAAAG-----	27358
QY	481	PheSerArgArgLySGlyLySGlyGlnGlnValProSerPheArgProTyrLeuLeu	500
Db	27537	---TCTGCACCTCGGAGACAGAGGT---GCGTCTGCACCGCTATACCGACTTGGCTGCT	27590
QY	501	Thr-----LeuPheThrGlnSerSerTyrProThrLeuHisGln	514
Db	27591	ACGCAAGCCCTGCACACGAGGCGCTTCGCGCTCCAAAGCTGTCATCGCCCTTCATCCG	27650
QY	515	CysLeuArgValLeuLeuGlyLysSerArgGlnGln-----ArgPheAspPro	530
Db	27651	TCTGGCCGAAAGCGACACTTCATGCGACAGCGCGCCGCGAGACACCGGCGACGCGCTCGCCT	27710
QY	531	SerAlaSerLeuAspPheLeuTrrAlaCysL1enHisValProArgTLeuProGlnArg	550
Db	27711	CTTGGAAAGCGCTG-----CGTCGCGACAGCGCGCTCGCGCTCGCGCGCTCGC	27758
QY	551	AspGlnArgThrProGlnLysArgArgGlnGlnLeuValLeuArgValGlnGlyProGln	570
Db	27739	---CGTCGTACCCGACGCGCGCGCGCGACCCACCGCTGAAGAAGAGTAAAGGCGCTCG-	27814
QY	571	LeuLeuSerLeuValGlnLeuLeuLeuLeuAlaGlnAlaGlnThrArgSerGlnAspLysAsp	590
Db	27815	CTCAGCGCGCTCTGGGGTCTCGCGCTCGCGCTCGCGAGAGACAGACCCAGAGCGCGCTC	27874
QY	591	ThrAlaAlaCysSerLeuLeuLeuGlnAlaArgLeuProLeuLeuLeuSerCysCysGly	610
Db	27875	TGCTGCTCGTCACTCGACGACAGACAGCGCGCTCCAGACGCGCTCGTGGCGCGCTCG	27934
QY	611	AspAspGluSerValArgLyValThrGlnHisLeuSerGlyCysL1eGlnGlnTrrGly	630
Db	27935	ACGCAAGAGAGCGACAGATCGCGCTCGCGACGCGAAACCCCTCGTCCAA-----GGC	27988
QY	631	AspSerValLeuGlyArgArgCysArgAspLeuLeuGlnLeuTyrLeuGlnArgPro	650
Db	27989	TCTCAAGCTGCCCCACAGCGCGCCACAGG-----ACACAGCGCTCCG	28027
QY	651	GluLeuArgValProValProGluValLeuHisSerGlnGlyAlaAlaSerSerSer	670
Db	28028	CGGCAAGGCTCGGAGGACACCTGC-CTATCAGCGGAGGACGCGGACGCTCGGCGGCTCG	28086
QY	671	ValCysLysLeuAspDlyLeuLeuLeuHisArgPheLeuLeu-----	685
Db	28087	GTCGCGCGCGCGCTGCTAAACCAACGACGCGCAAGACACTGCTGCTCACTCGCGCGAC	28148
QY	686	-----AlaAspThrSerAspSerArgLysGlnHisAsnArgGlyAla	699
Db	28147	GCGCGAGCGCTCGGCGGTCTATGCTCTTGCAGACGAG-CTCGAGACTGCGGGGCT	28203
QY	700	AspAlaSerMetAlaCysArgLyLeuAlaValAlaHisProLeuLeu--LeuLeuArg	718
Db	28204	TTCGTCACCTCGCGCGCGAGCGAGCGGCGATCCAGCGCGCTGTAAAGGACTTCTGAT	28263
QY	719	HisLeuProMet-----LeuAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPhe	735
Db	28264	AACATTCGACGAGCGCTCACCGCGGTGCGCGCGCTGTCGACAGCGCGCAAGCTCGACGCG	28333
QY	736	GlnGluPheArgGlnGlnHisHisLeuSerCysPheLeuHisValLeuGlnGlyLeuGlnGly	755
Db	28324	GAT-----CTGCTCGGCGCCATGAC	28344
QY	756	LeuLeuGln--ProHisValPheArgSerGlnHisGlnGlyAlaLeuTrrAspCysLeu	774
Db	28345	CTCGAGCGGATCGACCGGCTCTTCCGCGCGCAAGATCATGATCGCGC---TGGCAC-----	28395
QY	775	LeuSerPheLeuArgLeuLeuLeuHisGlnTyrArgLysSerSerArgHisLeuAlaPhe	794
Db	28386	-----TGCATGACGCTCACCAAGATTAACGCGCTTGCAGCGCTTCG	28434

```

      QY      795  Ile-----AspLysPheValGlnPheIleHisLysTyrIleThrTyrsAlaProA 812
          |||
      Db 28435 ATCTCTTCGTCCGTCGGCGCGGCCTCCTCGGCACTAGATCATCTGCCAACTACGCGCT 28496
          |||
      QY      813  AlaIleSerPheLeuGlnLysHisAlaasp-----ProLeuHis 825
          |||
      Db 28495 GCGACCGCCTTCCTCGATCCGCTTGCCGCAACCACCGCGCGCGCAAGGCGCTCCCTGCTCA 28554
          |||
      QY      826  AspLeuSerPhePheAspAsn---SerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeu 844
          |||
      Db 28555 TCGCTGCGCTGAAGCACACTGGCGCGCGGCGCACCGCAATACAGACAGACGTCAAGCGCCG 28612
          |||
      QY      845  SerLeuProSerArgAspAsparGthrAspArg---GlyLeuAspLeuLysGluGlu 863
          |||
      Db 28615 GCGCGCCCT-----CGCATGGAGCGCGCGCGGCTTCCTCGACCTGTGAGGAG 28664
          |||
      QY      864  GluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThr---- 882
          |||
      Db 28663 AGG-----CTCGCCCTTCGTGATGGCGGCGCTCTTCGACACGAGACCGCC 28707
          |||
      QY      883  -----AlaAlaGluMetAlaPro 888
          |||
      Db 28708 CTGGTCCCCGCGCGCTTCGACTTAGACGGCGCTCAGGCGGCAACGCCGCGACGCTCCCCCG 28766
          |||
      QY      889  TyrMetLysArgLeuSerArgLysGlnThrVal----- 899
          |||
      Db 28768 TTGTTCCAAAGCTCTGCTCGCGCTCGACACCTAACCGAAGCGCGCGCAACACCGCCCGAG 28827
          |||
      QY      900  ---GluAspLeuLeuGlnValLeuSerAspIleAspGluMetSerArgArgProGlu 918
          |||
      Db 28828 GCCTCGTGCCTTACAGAGCGCTCTCTAGCCCTCCCGCGCGCGGACGCAAGCGACGCT---GCC 28884
          |||
      QY      919  IleLeuSerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAla 934
          |||
      Db 28885 CTGCTCGATCTCATCCGACCGACGAGCGCGCGCTTCGCGCTCGCC 28932
          |||
RESULT 10
US-08-457-335A-6
Sequence 6, Application US/08457335A
Patent No. 5723759
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cliba-Gelgy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457, 335A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457, 205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258, 261

```

FILING DATE: 08-Jun-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James Scott  
 REGISTRATION NUMBER: 36,129  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-457-335A-6

Alignment Scores:  
 Pred. No.: 0.006 Length: 28958  
 Score: 150.50 Matches: 255  
 Percent Similarity: 34.668 Conservative: 111  
 Best Local Similarity: 24.15% Mismatches: 405  
 Query Match: 2.87% Indels: 292  
 DB: 1 Gaps: 57

US-09-929-769-7 (1-1029) x US-08-457-335A-6 (1-28958)

QY 14 ThrLeuGlyProProArgAlaAspSerGluPheGlnAlaLeuLysP----- 30  
 DB 26222 ACCCTTCGTCGAGGTACAGCCCATCCGTCACGCTCGCCCTCGCGAGACTGC 26280  
 QY 31 IleTrpPheProGluGluLysProLeuPro-----ThAlaPheLeuAlaSerPrrSer 48  
 DB 26281 CCCTTCCTGTCGAGGTACAGCCCATCCGTCACGCTCGCCCTCGCGAGACTGC 26340  
 QY 49 GluGlnAlaLeuLeuLeuProAspTrpLeu---LysLeuArgMetIleArgSerGluAla 67  
 DB 26341 GAGCGCTACACCGCTCGATCCGTCGTCGCTCGATTCGACGCGACGACGACGACGCTC 26400  
 QY 68 LeuAlaGluValAlaAlaLeuGlnAspLeuGluProGlnGlnLeuLeuPheVal 87  
 DB 26401 GCCCGCTGCTC---CTCTCGGCGGAGCTCTACCCGAGGCTCGCGCTCGACTCG 26457  
 QY 88 GlnSerPheGlyIleProValSerSerMetSerLysLeuLeuGlnPheLeuAspGlnAla 107  
 DB 26458 AACGCTTCCTGCGCGCTTCGCTCGCCGACAGGCTCCCTCCGACCTAC----- 26508  
 QY 108 ValAlaHisAspProGlnThrLeuGlnGlnAsnIleMetAspLysAsnTrpMetAlaHis 127  
 DB 26509 -----CCCTTCACAGCGCGCTTCGCTCGAGCCCTC---ACGGCGAC 26553  
 QY 128 LeuValGluValGlnHisGluArgGlyAlaSerGlyGlyGlnThrPhe-----HisSer 145  
 DB 26554 GCTGCGAGCTC-----GCCCTCGAGGCTGACCTCGGCGGACGCCG 26598  
 QY 146 LeuLeuThrAlaSerLeuProProArgArgAsp----- 156  
 DB 26599 CTGCTCGGCGCGCGCTCGC---CCTGCGACGCGGATGCGCTTTGCTTCACAGAGCGCT 26657  
 QY 157 -----SerThr 158  
 DB 26658 CTCCTCGCAGAGACCCGTCGTCGACAGACGTCGTCCTTCGGCATACCGCTGCTCTGC 26717  
 QY 159 GluAlaProLysProLysSerSerProGlu---GlnProIleGlyGlnGlyArgIleArg 177  
 DB 26718 CAGGGCGCCCTCTCGACTCGCCCTGCATGCTTCGTCGCGCTCGACACCGCT 26777  
 QY 178 ValGlyThrGlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGln 197  
 DB 26778 CGAAGACGTCACGCTCGA----- 26795

QY 198 IlePheProLeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeu 217  
 DB 26796 -----CCCCCCCCCTGCTCTCCCATGCGAGGGCGCGCTCTCTCA----- 26837  
 QY 218 AlaLeuGlnGlnAlaLeuGlyGlnGluLeuAlaArgValAlaGlnGly---SerProGlu 236  
 DB 26838 GATCTCGTCGCGCGCGCGGACGCTG---CTGAC-----GAAAGGCGCTCTCGCTTC 26887  
 QY 237 ValProGlyIleThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHis 256  
 DB 26888 ATAGCGGCGCCACGACGCGCTTC-----AGATGCGCTGACTCGCCACGCCA--- 26938  
 QY 257 GlyGlyAlaLeuValMetSerMetHisArgSerHisPheLeuAlaCys----- 272  
 DB 26939 -----GCGGCTCTCTCGCGAAGCTAGCCGCTCC---CATTGCTTGATGCTCCGCGATGCG 26994  
 QY 273 ---ProLeuLeuArgGlnLeuCysGlnTrpGlnArgCysValProGlnAspThr---Gly 290  
 DB 26995 CCCCCC-----TCGGGCGCCACCCAGGTGACACCCAGGT 27030  
 QY 291 PheSerSerLeuPheLeuLysValLeuLeuGlnMetLeuGlnTrpLeuAspSerProGly 310  
 DB 27031 TTCTACGACGCC-----CTCGAGACGCTGG 27057  
 QY 311 ValGluGlyPro-----LeuArgAlaGlnLeuArgMet 322  
 DB 27058 CTTCGCTATGCGCCGAGTTCACAGGCTCCGCGCGCGCTTCACACGCGCGACGAGCT 27117  
 QY 323 LeuAlaSerGlnAlaSerAlaGlyArg-----LeuSerAsp 335  
 DB 27118 CTTCGCGGAGGACCACTCCGCGACCGCCGACGAGGAGCGCGCTGTTTCCCTCA 27177  
 QY 336 ValArgGlyGlyLeuLeuArgLeuAlaGlnAlaLeuAlaPheArg----- 350  
 DB 27178 CCCCCCTGCTCGACAGGCTTCGACAGGCTCTGCTTGTAGACAGGACGACCAAGGC 27237  
 QY 351 ---GlnAspLeuGluValAlaSerSerThrValArgAlaValAlaLeaThrLeuArgSer 369  
 DB 27238 CTTCAGAGATGCCCTTCTCGGAGCGAGT-----ATCGTGGCTCGCGTGC 27285  
 QY 370 GlyGluGlnCysSerValGluProAsp-----LeuIleSerLysValLeuGlnGly 386  
 DB 27286 -----GAGCCACACCTCGCGCTGCTTCACCGCTCGAGAGCGG 27327  
 QY 387 LeuIleGluValArgSerProHisLeuGlnGluLeuLeuThrAlaPheSerAlaThr 406  
 DB 27328 AATCTCGGCGCTGCTC-----TCCTGCGCG 27354  
 QY 407 AlaAspAlaAlaSerProPheProAlaCysLys-----ProValValAlaValSerSer 424  
 DB 27355 ACGCCAGAGCGCAACCATCGCTCGTGACAGCGCTCGGCATGCGCG----- 27402  
 QY 425 LeuLeuLeuGlnGluGluProLeuAlaGlyGlyLysProGlyAlaAspGlySer 444  
 DB 27403 -----CCGGCTCGCGAGACGCTCGCAGACCCGGAGCG-----TCCAC 27444  
 QY 445 LeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuGlnMetLeu 464  
 DB 27445 CTGATGCGCTC-----TTCGATCATGACTGAGCGAGCTGCA- 27482  
 QY 465 AspProGluValAlaSer-----SerCysProAspLeuGlnLeuArgLeuLeu 480  
 DB 27483 AACGCCACACTCACGCCCATCGCCCGGAGCGGCTCTCTCGGACAGAAAG----- 27536  
 QY 481 PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProTrpLeuLeu 500  
 DB 27537 ---TTCGACCTCGGAGCAGAGT---GCCCTCGACCGCTATACCGACCTCTCTCTCT 27590  
 QY 501 Thr-----LeuPheThrHisGlnSerSerTrpProThrLeuHisGln 514  
 DB 27591 ACGCAGCGCGCTCGACAGGCGCTTCGCTCAACGCTTCATGCGCCCTTCATCGC 27650  
 QY 515 CysIleArgValLeuLeuGlyLysSerArgGluGln-----ArgPheAspPro 530



```

Db 27651 TCTGCCGAGGCGACCTCATCCGAGCGCCGAGACCCGCGCCGCTGCGCTC 27710
OY 531 SerAlaSerLeuSerPheLeuThrAlaCysIleHisValProArgIleTrpGlnGlyArg 550
Db 27711 CTTCGACAGCTG-----GCTCGCCGAGACCGCGCTGCTCTCGCGCGCTCGC 27758
OY 551 AspGlnArgThrProGlnIleArgArgGlnGlnLeuValLeuArgValGlnGlyProGln 570
Db 27759 ---CTGCGTACCCGAGCGCGCGCTCGCCACCCGCTGAGAAAGAGCTCAAGGCGCTCG- 27814
OY 571 LeuIleSerLeuValGlnLeuIleLeuAlaGlnAlaGluThrArgSerGlnAspGlyAsp 590
Db 27815 CTGACGCGCGCTCTGCGGCTCGCTCGCTCGCGCGAGACGAGACCGAGCGCGCTC 27874
OY 591 ThrAlaAlaCysSerLeuIleGlnAlaArgLeuProLeuLeuSerCysCysGly 610
Db 27875 TCTCTCTCGTGCAGCTCGACGACGAGCGCGCTCGCGCGCGCTCTGCGCGCGCTCG 27934
OY 611 AspArgGlnSerValArgGlyValThrGlnIleLeuSerGlyCysIleGlnGlnTrpGly 630
Db 27935 ACCGACAGAGAGAGATGCGCTCGCGACGAGGAGAAAGCGCTGTCGAA-----GCG 27988
OY 631 AspSerValLeuGlnArgArgCysArgAspLeuLeuLeuLeuIleuTrpLeuGlnArgPro 650
Db 27989 TCTGACGCGCTGCGCGCGCGCGCGCGCG-----ACACAGCGCTCGC 28027
OY 651 GlnLeuArgValProValProGlnValLeuLeuHisSerGlnGlnAlaAlaSerSer 670
Db 28028 CCGCAGGCGCTCGAGGAGCGCGTCTCATACAGGAGGAGCGACCGCGCGCGCGCTG 28086
OY 671 ValCysLeuLeuArgGlyLeuIleHisArgPheIleThrLeuLeu-----685
Db 28087 GTTGGCGCGCGCGCTCGTGTAAACACGAGCGCGACCGCTGCTCGCTCGCGCGCGCAG 28146
OY 686 -----AlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAla 699
Db 28147 GCGCGCAGCGCTCGCGGTGTATGCTTTCGGAAGCGAG---CTCGAGCTCTGGGGGCT 28203
OY 700 AspAlaSerMetAlaCysArgGlyLeuAlaValAlaHisProLeuLeu---LeuLeuArg 718
Db 28204 TCGGTACCGCTCGCGCGTGCAGCGCGCGAGTCCAGCGCGCTGTAAAGAGACTTGTGAT 28263
OY 719 HisLeuProMet-----IleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPhe 735
Db 28264 AACATTTCGAGCGCTGACCGCGGTGCGCGCTGTGATCCCGCGCTGCTGAGCGC 28323
OY 736 GlnGluPheArgGlnGlnAsnHisLeuSerCysPheLeuHisValLeuGlnLeuGln 755
Db 28324 GAT-----CTGCTCGCGCGCGCTGAGC 28344
OY 756 LeuLeuGln---ProHisValPheArgSerGlnHisGlnGlnIleLeuTrpAspCysLeu 774
Db 28345 CTGACGCGCGTGCAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 28395
OY 775 LeuSerPheIleArgLeuLeuLeuAsnTrpArgGlySerSerArgHisLeuAlaPhe 794
Db 28396 -----TTGCATCAGCTGACCGCGCGCGCGCGCGCGCGCGCTG 28434
OY 795 Ile-----AsnLysPheValGlnPheIleHisLysTrpIleThrTrpAsnAlaProAla 812
Db 28435 ATGCTCTTCTCGTCTGCGCGCGCGCGCTGCGCGAGCTCAGAGTCACTCACTACGCGCT 28494
OY 813 AlaIleSerPheLeuGlnIleHisAlaAsp-----ProLeuHis 825
Db 28495 GCGAGCGCGCTTCTGATGCGTGGCGACACCGCGCGCGCGCGCGCGCGCGCTG 28554
OY 826 AspLeuSerPheAspAsn---SerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeu 844
Db 28555 TCGCTGCGGTGAGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28614
OY 845 SerLeuProSerArgAspArgTrpAspArg---GlyLeuAspGlnGlnGlnGln 863

```

```

Db 28615 GCGCGCGCT-----CGCATGAGCGCGCGCGCTTCCCTCGACTCTGAGAG 28662
OY 864 GluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThr--- 882
Db 28663 AGG-----CTCGCGCTCTTGATGCGCGCGCTCTTCCGACCGAGACCGCGC 28707
OY 883 -----AlaAlaGlnMetAlaPro 888
Db 28708 CTGATCCCGCGCGCTGCACTTGAGCGCGCTGAGGCGGAGCGCGCGCGCGCGCG 28767
OY 889 TyrMetLysArgLeuSerArgGlyGlnThrVal-----899
Db 28768 TTGTTCCACGCTCTGCGCGCTGCGCGCGTACCGACGCGCGCGCGCGCGCGCGCG 28827
OY 900 ---GluAspLeuGlnGlnValLeuSerAspIleAspGlnMetSerArgArgProGln 918
Db 28828 GCGTCTGCTTACAGAGCGCGCTCTGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 28884
OY 919 IleLeuSerPhePheSerThrAsnLeuGlnArgLeuMetSerAla 934
Db 28885 CTGCTGATGTCATCCGACCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCG 28932

RESULT 11
US-08-729-214-6
Sequence 6, Application US/08729214
Patent No. 5817502
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Hammett, Phillip E.
APPLICANT: Van Pee, Karl-Helinz
APPLICANT: Kirner, Sabine
TITLE OF INVENTION: Genes for the synthesis of
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/729,214
FILING DATE: TBA
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,741
REFERENCE/DOCKET NUMBER: CGC 1506/CIPS
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-729-214-6

Alignment Scores:
Pred. No.: 0.006 Length: 28958
Score: 150.50 Matches: 255

```



Percent Similarity: 34.66% Conservative: 111  
 Best local Similarity: 24.15% Mismatches: 405  
 Query Match: 2.87% Indels: 292  
 DB: 1 Gaps: 57

US-09-929-769-7 (1-1029) x US-08-729-214-6 (1-28958)

QY 14 ThrLeuGlyProProArgAlaAspSerGluPheGlnAlaLeuLeuAsp----- 30  
 DB 26222 ACCTCCGGCAAAACCTCTCTGTTTC-TCGAGCGGAGACCGGCTCTGACATGGGCAAT 26280  
 QY 31 IleTrpPheProGluGluLysProLeuPro-----ThrAlaPheLeuValAspTrpSer 48  
 DB 26281 CGCTCTTCGTCGAGGTGAGCCGCCATGCCGCTGACAGCTGCGCTCCGCGAGACGTGC 26340  
 QY 49 GluGluAlaLeuLeuLeuProAspTrpLeu---LysLeuArgMetLeuArgSerGluVal 67  
 DB 26341 GAGCGCTACCGCTGATCCGCTGCTGCGCTCCATTCGACGCGAGACGACGACCTC 26400  
 QY 68 LeuArgLeuValAspAlaAlaLeuGlnAspLeuGluProGlnGlnLeuLeuPheVal 87  
 DB 26401 GCCCGCTGCTC---CTCTCGGGGAGAGCTCTACCGGAGGCTCGCGCTCCACTGG 26457  
 QY 88 GlnSerPheGlyIleProValSerSerMetSerLysLeuLeuGlnPheLeuAspGlnAla 107  
 DB 26458 AACGCTCTTCGCGCGCTGCTGCGCGAGGCTGCTCCGCTCCGACCTAC----- 26508  
 QY 108 ValAlaIleAspProGlnThrLeuGlnGlnAsnIleMetAspLysValAsnTrpMetAlaHis 127  
 DB 26509 -----CCCTCCCAACGCGAGCGCTTGTGCTGACGCGCTC---ACGCGCGAC 26553  
 QY 128 LeuValGluValGlnHisGluArgGlyAlaSerGlyGlyGlnThrPhe-----HisSer 145  
 DB 26554 GCTGCCGACGTC-----GCTCCGAGCGCTTACCTCGGCGGACCAACCGC 26598  
 QY 146 LeuLeuThrAlaSerLeuLeuProProArgArgAsp----- 156  
 DB 26599 CTGCTCGCGCGCGCGCGCTGCG---CCTCGCGAGCGCGATGGCTTGTCTTCACAGACGCGT 26657  
 QY 157 ----- 158  
 DB 26658 CTCCTCCGACGACGACCGCTGCTGCGAGACGACGCTCTTGGCGATACCTGCTCTGC 26717  
 QY 159 GluAlaProLysProLysSerSerProGlu---GlnProIleGlyGlnGlyArgIleArg 177  
 DB 26718 CAGGCGCGCGCTCTCGAGCTGCGCTGATGCTCCGCCATCTCGTCGCGCTGACACCGT 26777  
 QY 178 ValGlyThrGlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGln 197  
 DB 26778 CGAAGACGTCACGCTCGA----- 26935  
 QY 198 IlePheProLeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeu 217  
 DB 26796 -----CCCCCCCCCTGCTCTCCATCGCGAGCGCGCTCTCTCTCA-- 26837  
 QY 218 AlaLeuGlnGlnAlaLeuGlyGlnGlnLeuAlaArgValAlaGlnGly---SerProGlu 236  
 DB 26838 GATCTCCGTCGCGGCGCGGAGCGGTG-CTGAC-----GAAGCGGCTCTCTCGTTC 26887  
 QY 237 ValProGlyIleThrValAlaArgValLeuGlnAlaLeuAlaThrLeuLeuSerProHis 256  
 DB 26888 ATAGCGGCGCGACGACGCGCTTC-----AGATGGCGCCCTGAGCTCGCACGCGCA-- 26938  
 QY 257 GlyGlyAlaLeuValMetSerMetHisArgSerHisPheLeuAlaCys----- 272  
 DB 26939 ---GGGCTCTCTCGCGAAGCTACCGCTCC-CATTGCTTTCGATGCTCCGCGCATATGC 26994  
 QY 273 ---ProLeuLeuArgGlnLeuGlyGlnTrpGlnArgCysValProGlnAspThr---Gly 290  
 DB 26995 CCCCC-----TCGGGCGCGCACCCAGAGTGCACCAAGGT 27030  
 QY 291 PheSerSerLeuPheLeuLysValLeuLeuGlnMetLeuGlnTrpLeuAspSerProGly 310  
 DB ----- 311

DB 27031 TTCACGCGACG-----CTCGAGACGCGCTGGG 27057  
 QY 311 ValGluGlyGlyPro-----LeuArgAlaGlnLeuLeuArgMet 322  
 DB 27058 CTTCGCTTATGCGCCAGGTTCAGGCGCTCGCGCGCTGTACAGCGCGGAGAGCT 27117  
 QY 323 LeuAlaSerGlnAlaSerAlaGlyArg-----LeuSerAsp 335  
 DB 27118 CTCGCGCGAAGCGACGCTCCGCGAGCGCGCGCGCGCGCGCGCGCTTTGGCTCTCA 27177  
 QY 336 ValArgGlyGlyLeuLeuArgLeuAlaGlnAlaLeuAlaPheArg----- 350  
 DB 27178 CCCCCCGCTCGACAGCGCTTCGAGCGCTCGCTTTCAGAGACGACGCAAGGCG 27237  
 QY 351 ---GlnAspLeuGlnValValSerSerThrValAlaAlaValIleAlaThrLeuArgSer 369  
 DB 27238 CTTCAGAGATGCGCTTCTCTGAGCGGAGT-----ATGCTCGCTCGCGCTCG 27285  
 QY 370 GlyGluGlnCysSerValGluProAsp-----LeuIleSerLysValLeuGlnGly 386  
 DB 27286 -----GAGCACACCGCTCGCGCTGCGCTTCACCGCTCGAGAGGCG 27327  
 QY 387 LeuIleGluValArgSerProHisLeuGluGluLeuThrAlaPheSerAlaThr 406  
 DB 27328 AATCTCGCGCTCGCTC-----TCCTCGCGG 27354  
 QY 407 AlaAspAlaAlaSerProPheProAlaCysLys-----ProValValValValSerSer 424  
 DB 27355 ACGCAGAGCGAGCAACCATCGCTCGGTCAAGCGCTCGCGCTCGCGCG----- 27402  
 QY 425 LeuLeuLeuGlnGluGluProLeuAlaGlyGlyLysProGlyAlaAspGlyLys 444  
 DB 27403 -----CCGCGTCCGCGAGACGACTCCGAGACCGCGGAGCG-----TCCAC 27444  
 QY 445 LeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuGluMetLeu 464  
 DB 27445 CTTCGAGCGCTC-----TTCCGATGACGACGAGCGAGCTGCA- 27482  
 QY 465 AspProGluValValSer-----SerCysProAspLeuGlnLeuArgLeuLeu 480  
 DB 27483 AAGCCCGACCTACGCGCCATCGCGCGCGCGCGCGCTCTCTCGGACGAGAAAG----- 27536  
 QY 481 PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProTrpLeuLeu 500  
 DB 27537 ---TTCGACGTCGCGAGACAGGCT---GCTCGACGCGCTATACGAGCTGCTGCTCT 27590  
 QY 501 Thr-----LeuPheThrHisGlnSerSerTrpProThrLeuHisGln 514  
 DB 27591 ACGCAGCGCTCGACGAGCGCGCTTCCCTCCAAACCTCGCATCGCGCTTCATCGC 27650  
 QY 515 CysIleArgValLeuLeuLysSerArgGluGln-----ArgPheAspPro 530  
 DB 27651 TCTGCCCGAAGCGACCTCATGCGAGCGCGCGCGAGACCGCGCGCGCTCGCGCT 27710  
 QY 531 SerAlaSerLeuAspPheLeuThrAlaCysIleHisValProArgIleTrpGlnArg 550  
 DB 27711 CTTCGAGCGCTG-----GCTCGCGAGCGAGCGCGCTCGCTCGCGCGCTCGC 27758  
 QY 551 AspGlnArgThrProGlnLysArgArgGluGluLeuValLeuArgValGlnGlyProGlu 570  
 DB 27759 ---CTCTCGACCGCGCGCGCTGCGCGCGCGCGCGCGCGCTGCGCGCTCGC- 27814  
 QY 571 LeuIleSerLeuValGluLeuIleLeuAlaGlnAlaGlnThrArgSerTrpAspGlyAsp 590  
 DB 27815 CTCAGCGCGCTCTCGGGGTCTCGCTCGCTCGCGCGCGAGCGAGCGACCGCGCGCTC 27874  
 QY 591 ThrAlaAlaSerSerLeuIleGlnAlaArgLeuProLeuLeuLeuSerCysCysGly 610  
 DB 27875 TCGTCTCTGTCACCTCGACGAGCGAGCGCTCCAGACCGCGCTCGCGCGCGCTCG 27934  
 QY 611 AspAspGluSerValArgLysValThrGluHisLeuSerGlyCysIleGlnGlnTrpGly 630  
 DB 27935 ACGCAAGAGAGCGAGATCGCGCTCCGCAAGCGCAACCGCTCGTTCCA-----GCG 27988



[illegible]

Db	27238	CTTCAGAGATGCCCTTCTCTGTGGACGGAGT-----ATGCGTGGCTCCGGCTCG	27285
QY	370	glyIugInCysSerValGluProAsp-----LeuIleSerIysValLeuGIngly	386
Db	27286	-----GAGCCACACACCTTCGGGGTGGTTTCCACGCTCTGAGGGCG	27327
QY	387	LeuIleGluValArgSerProHisIleuGluIleuLeuThrAlaPheSerAlaThr	406
Db	27328	AAATCTCGCGGCTCGCTCC-----TCTCTCGCGG	27354
QY	407	AlaAspAlaIleSerProPheProAlaCysIys-----ProValValValSerSer	424
Db	27355	ACGCCAGAGGCGAACCCATCGCTCGGTGCACAGCCCTGCCATGCGCG-----	27402
QY	425	LeuLeuLeuGIngluGluIleProLeuAlaGlyIysProGluValAspIlyGlySer	444
Db	27403	-----CCGCGTCCGCCGAGACAGACTCCGAGACCCGGAGAGC-----TCCAC	27444
QY	445	IeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuIleu	464
Db	27445	CTCATGTCCTCC-----TTCCGATGACTGGAGGAGCTGCA-27482	
QY	465	AspProGluValValSer-----SerCysProAspLeuGInleuArgLeu	480
Db	27483	AAGCCCATCTACCGCCCATCGCCCGAGCGGTGCCCTCTCGGACAGAGG-----	27536
QY	481	PheSerArgIysGlyIysGlyIleGluIleValProSerPheArgProTrpLeu	500
Db	27537	-----TCTCGACCTCGGACCAAGGT-----GCTCTCGACCCCTTACCAGACTTGTCTCT	27590
QY	501	Thr-----LeuPheThrHisGlnSerSerTrpProThrLeuHisGln	514
Db	27591	ACGAGAGCCCTTCGACACAGCGGCTTCGCCCTCCAAACCTGTCATCGCCCTTATCGC	27650
QY	515	CysIleArgValLeuLeuGlyIysSerArgGluGln-----ArgPheAspPro	530
Db	27651	TCGCCCCGAGGAGCACTTCATCGACAGCCCGCGAGACCAACCGCGACGCTCGCT	27710
QY	531	SerAlaSerLeuAspPheLeuTrpAlaCysIleHisValProArgIleTrpGlnArg	550
Db	27711	CTTCGACGCTG-----GCTCGCCAGACAGGAGCTCGCTCTCGCGCTCGC	27758
QY	551	AspGlnArgThrProGlnIysArgArgGluIleuValLeuArgValGlnIlyProGlu	570
Db	27759	-----CCTCGTCACCCGACGCGCGCTCGCGCCACCGCTGAAGAGAGCTCGC-	27814
QY	571	LeuIleSerLeuValGluIleuLeuAlaGluAlaGluThrArgSerGlnAspIleAsp	590
Db	27815	CTACAGCGGCTTCGGGGTCTCGCTCGCTCGCGCAGAGACCAACCGAGCGGCTCT	27874
QY	591	ThrAlaIleCysSerLeuIleGlnAlaArgLeuProLeuLeuLeuSerCysCysGly	610
Db	27875	TCGTCTCTCGACCTCGACAGACAGACAGGAGGCTCCAGACAGCCGCTCGCGCGCTCG	27934
QY	611	AspAspGluSerValArgIysValThrGlnHisLeuSerGlyCysIleGlnIleTrpGly	630
Db	27935	ACGCAAGAGAGCAGAGATCGGCTTCGCCAAGCGCAAAACCTCTGTTCA-----GCG	27988
QY	631	AspSerValLeuGlyArgArgCysArgAspLeuLeuGlnLeuIleuTrpLeuIleArgPro	650
Db	27989	TCGACGCTTCGCCAGGCGGCCACGG-----ACACAGCGTCC	28027
QY	651	GluLeuArgValProValProGluValLeuLeuHisSerGluGlyAlaAlaSerSerSer	670
Db	28028	CCGCAAGGCTCGGAGGACCGCTC-CTCATCAGCGGAGGACCGGACGCTCGCGCCCTG	28086
QY	671	ValCysIysLeuAspIlyLeuIleHisArgPheIleThrLeu-----	685
Db	28087	GTGCGGCGCGCTGTCTAAACACAGACGCGCAAGACACTGTCTCTCACTCGCGCCAG	28146
QY	686	-----AlaAspThrAspSerArgAlaLeuGlnAsnArgGlyAla	699

```

Db 28147 GCGCGAGCGCTCGGCTGTATGCTTCGCGAGAG--CTCGAGCTTGGGGCT 28203
QY 700 AspAlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeu---LeuLeuArg 718
Db 28204 TCGGTACACCTCGCGGTGCGAGCTGGCGGATCGACAGCGCTTAAGAGCATTTGAT 28263
QY 719 HisLeuProMet-----IleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPh 735
Db 28264 AACATTCCGAGCGCTACCGCGGTGCGCGCTGTGATCGCGACCGCTCTCGAGCG 28323
QY 736 GluGluPheArgGlnGlnAsnHisLeuSerCysPheLeuHisValLeuGluLeuGlu 755
Db 28324 GAT-----CTGCTCGCGCGCATAGC 28344
QY 756 LeuLeuGln---ProHisValPheArgSerGlnHisGlnValAlaLeuTrpAspCysLeu 774
Db 28345 CTGACGCGGATCGACGCGCTTCGCGCGCGCATGATCGATCGCGC---TGGCAC--- 28395
QY 775 LeuSerPheIleArgLeuLeuLeuAsnTyrArgLysSerSerArgHisIleAlaIlePhe 794
Db 28396 -----TTGCATCAGCTGACCCAGATAGCCCTTGGCGCTTC 28434
QY 795 Ile-----AsnLysPheValGlnPheIleHisLysTyrIleThrTyrAsnAlaProAla 812
Db 28435 ATCCTCTTCGTCGTCGCGCGCGCGCTCTGCGAGCTCAGCTCAGCTCAGCTCAGCGCT 28494
QY 813 AlaIleSerPheLeuGlnLysHisAlaAsp-----ProLeuHis 825
Db 28495 GCGAGCGCGCTTCCTGATGCGGTGCGCACCGCGCGCGCGCGCGCGCGCTGCTGCTCA 28554
QY 826 AspLeuSerPheAsn---SerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeu 844
Db 28555 TCGCTGCGGTGAGCGACCTGCGCGCGCGCGCGCGCATGACAGACGACGACGCGCGCC 28614
QY 845 SerLeuProSerArgAspAspArgThrAspArg---GlyLeuAspGlnGlnGluGlu 863
Db 28615 GCGCGCGCT-----CGCATGAGACGCGCGCGCTTCCCTGAGCTGAGAGAG 28662
QY 864 GluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThr--- 882
Db 28663 AGG-----CTCGCGCTTCGATGCGCGCTTCGCGCGCGCGCGCGCGCGCGCC 28707
QY 883 -----AlaAlaGluMetAlaPro 888
Db 28708 CTGCTCCCCCGCGCTTCGACTTGAGCGCGCTAGGCGCGCGCGCGCGCGCGCGCG 28767
QY 889 TyrMetLysArgLeuSerArgGlyGlnThrVal-----899
Db 28768 TTGTTCCAGCTCTCGCGCGCGCTGCGACCGCTACGAGCGCGCGCGCGCGCGCGCG 28827
QY 900 ---GluAspLeuGlnValLeuSerAspIleAspGlnMetSerArgArgProGlu 918
Db 28828 GCGCTGCTGCTTACAGAGCGCGCTCTGAGCGCTCGCGCGCGCGCGCGCGCGCGCG 28884
QY 919 IleLeuSerPheSerThrAsnLeuGlnArgLeuMetSerAla 934
Db 28885 CTGCTGATGTATCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28932

```

```

; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,118
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528
; FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ. ID NO. 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4031 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
;
US-08-993-118-1
;
Alignment Scores:
; Pred. No.: 0.000359 Length: 4031
; Score: 147.00 Matches: 203
; Percent Similarity: 33.80% Conservative: 113
; Best Local Similarity: 21.71% Mismatches: 312
; Query Match: 2.80% Indels: 310
; DB: 2 Gaps: 38
;
US-09-929-769-7 (1-1029) x US-08-993-118-1 (1-4031)
;
QY 86 PheValGlnSer---PheGlyIleProValSerSerMetSerLys-----LeuLeuGln 102
Db 728 TTTTCCCGAGTCTCTCTCCAGATTCCTGTGAGCTCTCTCTCTCTCTCTCTCTCTCT 787
QY 103 PheLeuAspGlnAlaValAlaHisAsp-----ProGlnThrLeu 115
Db 788 CTTTCCAGAGTTCCCTCGAGAGACTGAGAGACTTTTGAGGGTTTCCCGAGTCTCT 847
QY 116 GluGlnAsnIleMetAspLysAsnTyrMetAlaHisLeuValGlnValGlnHisGluArg 135
Db 848 CTCAGATTCCTATGACCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 907
QY 136 GlyAlaSerGlyGlnThrPheHisSerLeu-----146
Db 908 CCTGAGAGTGTCTAAAGTACTTTTGAGGGTTTCCCGAGTCTCTCTCTCTCTCTCTCT 967
QY 147 -----LeuThrAlaSerLeuProProArgArgAsp 156
Db 968 AGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1027
QY 157 SerThrGluAlaProLysProLysSerSerProGlnGlnProIleGlyGlnArgIle 176
Db 1028 AGTACTTTTGAGGGTTTCCCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1081
QY 177 ArgValGlyThrGlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeu 196
Db 1082 TTCTCTCTACTTATGAGTATTTTCCAGAGTTTCTCTGAGAGTGTCTAAAGTACTTT 1141
QY 197 GlnIlePheProLeuSerPro-----AspProArgTrpGlnSer 209
Db 1142 GAGGGTTTCCCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1201

```

ADDRESS: 805 Third Avenue  
 STREET: 805 Third Avenue



Db 2792 GCGCGTTCTCTCTCAATATCAAGTACAGCACCCTATCAAAAGCAGATGCTG 2851  
Qy 888 -----ProlymethylsargLeuSerArgly 896  
Db 2852 ACGATGTCATCAGACAGTACAGCGGCTACTTCTGTGATCTTCAGAAAGCCCGT--- 2908  
Qy 897 GlnThrValGlnAspLeuLeuGluValLeuSerAspIle 909  
Db 2909 -----GAGTTTCATAGAGTACTTTTGGCATT 2935

RESULT 14  
US-08-845-528C-1  
Sequence 1, Application US/08845528C  
Patent No. 6027924

GENERAL INFORMATION:  
APPLICANT: LUCAS, Sophie;  
APPLICANT: DE SMET, Charles;  
APPLICANT: BOON-FALLEUR, Thierry;  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR  
TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pelfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/845,528C  
FILING DATE: April 25, 1997  
CLASSIFICATION: 4335

ATTORNEY/AGENT INFORMATION:  
NAME: Mary Anne Schofield  
REGISTRATION NUMBER: 36,669  
REFERENCE/DOCKET NUMBER: LUD 5455  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4031 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double-stranded  
TOPOLOGY: linear  
US-08-845-528C-1

Alignment Scores:  
Pred. No.: 0.000359 Length: 4031  
Score: 147.00 Matches: 203  
Percent Similarity: 33.80% Conservative: 113  
Best Local Similarity: 21.71% Mismatches: 312  
Query Match: 2.80% Indels: 310  
DB: 3 Gaps: 38

US-09-929-769-7 (1-1029) x US-08-845-528C-1 (1-4031)

Qy 86 PheValGlnSer---PheGlyIleProValSerSerMetSerLys-----LeuLeuGln 102  
Db 728 TTTCCCAAGTCTCTCTCCAGATTCTGTGAGCTCTCTCTCCCACTTATTAGAGT 787  
Qy 103 PheLeuAspGlnAlaValAlaHisAsp-----ProGlnThrLeu 115  
Db 788 CTTTTCACAGTCCCTCAGAGACTGAGTCTTTGAGGGTCTTCTCCCAAGTCTCT 847  
Qy 116 GlnGlnAsnIleMetAspLysAsnIleMetAlaHisLeuValGlnIleHisGluArg 135

Db 848 CTCACATTCTCTTGACCT 907  
Qy 136 GlyAlaSerGlyGlyGlnThrPheHisSerLeu----- 146  
Db 908 CCTGAGAGTCTCTCAAGTACTTTTGAAGGTTTCCCAAGTCTCTCTCCAGATTCTGG 967  
Qy 147 -----LeuThrAlaSerLeuProProAlaGlyAsp 156  
Db 968 AGCCCT 1027  
Qy 157 SerThrGlnAlaProLysProLysSerSerProGlnIleProIleGlyGlnIleArgIle 176  
Db 1028 AGTACTTTTGAAGGTTTCTCCCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1081  
Qy 177 ArgValGlyThrGlnLeuArgValLeuGlyProGlnAspAspLeuAlaGlyMetPheLeu 196  
Db 1082 TTCT 1141  
Qy 197 GlnIlePheProLeuSerPro-----AspProArgThrGlnSer----- 209  
Db 1142 GAGGTTTCTCCCAAGTCT 1201  
Qy 210 -----SerSerProArg----- 213  
Db 1202 CTGAGTCTTTTCCAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1261  
Qy 214 ---ProValAlaLeuAlaLeuGlnGlnAlaLeuGlyGlnIleLeuAlaArgValAlaGln 232  
Db 1262 TCT 1321  
Qy 233 GlySerProGlnValProGlyIleThrValArgValLeuGlnAlaLeuAlaThrLeuLe 252  
Db 1322 AGTTTCT 1373  
Qy 252 userSerProHisGlyGlyAlaLeuAlaMetSerMetHisArgSerHisPheLeuLacy 272  
Db 1374 -----TTCAGATCTCTGTGAGC 1390  
Qy 272 sProLeu-----LeuArgGlnLeuGlyCysG 280  
Db 1391 TCT 1450  
Qy 280 nIlyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuLysValLeuL 300  
Db 1451 ACTTTGAGGGTTTCT 1501  
Qy 300 euGlnMetLeuGlnThrPheLeuAspSer---ProGlyValGlnGlyGlyProLeuArgAlaG 319  
Db 1502 TCT 1560  
Qy 319 ILeuArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValAlaArgGlyG 339  
Db 1561 TTTTGAAGGTTTCTCCCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1601  
Qy 339 ILeuLeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValAlaSerSerT 359  
Db 1602 -----GTCTCTCTGAGGGGAGAAATCCCATTTCTCTCTCTCTCTCTCTCTCT 1638  
Qy 359 hrValArgAlaValIleAlaThrLeuArgSerGlyGlnGlyCysSerValGlnProAspL 379  
Db 1639 GATTGTTCCCAAGTCT 1677  
Qy 379 euIleSerLysValLeuGlnGlyLeuIleGlnValArgSerProHisLeuGlnIleLeu 399  
Db 1678 TCT 1728  
Qy 399 euThrAlaPhePheSerAla-ThrAlaAspAlaAlaSerProPheProAlaCysLysPro 418  
Db 1729 TCACTACTTCT 1770  
Qy 419 ValValAlaValSerSerLeuLeuGlnGlnGlnIleGlnIleProLeuAlaGlyGlyLysPro 438  
Db 1771 -----GTCT 1815









QY 578 eleuAlaGluValuThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleG1 598  
Db 2300 G-----CCTCTAG 2308  
QY 598 nAlaArgLeuProLeuLeuLeuSerCysCysGlyAspAspGluSerValArgGlySva 618  
Db 2309 GGGAGGAGATCCCTGCTCTCC-----AAATTCCTCAGAGTCT 2350  
QY 618 lThGlnHisLeuSerGlyCysIleGlnGlnTProGlyAspSerValLeuGlyArgArgCy 638  
Db 2351 CTTCAGGAGAGAGACTCCCTGT----- 2372  
QY 638 sArgAspLeuLeuGlnLeuGlnTyrLeuGlnArgProGluLeuArgValProValProG1 658  
Db 2373 -----CTCTCTCCATTTCTTCAGA----- 2393  
QY 658 uValLeuLeuHisSerGlnGlyAlaAlaSerSerSerValCysIlySLeuAspGlyLeuI1 678  
Db 2394 -GTCTCTCTGAGTGG-----AGGACTCCCTCTCTCT 2425  
QY 678 eHisArgPheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGluAsnArgI 698  
Db 2426 CTCCACTTTCCTCAGTTTCTCTCTC-----AGGGG 2455  
QY 698 yAlaAspAlaSerMetAlaCysArgGlyLeuAlaValAlaHisProLeuLeuLeuAr 718  
Db 2456 GAGCACTTCCAGTCTCTCTCTCCAGAGTCTGTGAGTATCTGCTCTCCCTCTCTTTG 2515  
QY 718 gHisLeuProMetIleAla-----AlaLeuLeuHisGlyArgThrHisLe 733  
Db 2516 AGTCTTCCCTCAGAGTTTCTCTGAGAGTCTCAGAGTCTCCTGAGGGGCTCTCAGTCT 2575  
QY 733 uAsnPheGlnIuPheArgGlnGlnAsnHisLeuSerCysPheLeuHisValLeuGlyLe 753  
Db 2576 CCTCTCCAGAGAC-----CTGTACAGTCC-TCCTTCTCTACACTTT 2616  
QY 753 uLeuGluLeuGlnProHisValPheArgSerGlnHis----- 766  
Db 2617 AGCGAGTCTTCTCCA-----AGTCCCATGAGAGTCTCTCAGAGTCTCTC 2661  
QY 767 -----GlnGlyAlaLeuTyrAspCysLeuLeuSerPheIleArgLeuLeuLe 782  
Db 2662 TGAGGGGCTGCTCCCTCTCTCCAGAGTCTGAGTCTCTCCCTCTCTCAGTCT 2721  
QY 782 uAsnTyrArgIysSerSerArgHisLeuAlaAlaPheIleAsnIysPheValGlnPheI1 802  
Db 2722 ATCGAGTCTTCCAGAGTCTCTGTGAGTCTCTCCCTCTCTCCTCATCGAGTCT 2781  
QY 802 eHisIys-----TyrIleThrTyrAsnAlaProAl 812  
Db 2782 TTCCAGAGTTCCTGAGAGTCTCTCCAGAGTCTGTGATCTCTCTCTCTCTCCAC 2841  
QY 812 aAlaIleSer---PheLeuGlnIysHisAlaAspProLeuHisAspLeuSerPheAspAs 831  
Db 2842 TTCATTGAGCCCATTCAGTGAAGAGTCCAGAGCCCATAGATGAATATACAAAGTTCTC 2901  
QY 831 nSerAspLeuValMetLeuIysSerLeuLeuAlaGlyLeuSerLeuProSerArgAspAs 851  
Db 2902 AGACACCTTGTGAGAGTGTATCTCTGACAGACAGCATCTTGATAGAGAGCGAGCC 2961  
QY 851 PArgThrAspArgIlyLeuAspGlnGluGlyLeuGlnIuSerSerAlaGlySerLeuPr 871  
Db 2962 CTGTGTCACCTATACACTGATGAAGAGTGGAGAG-----TTGGCGCGGTTTCT 3012  
QY 871 oLeuValSerValSerLeuPheThrProLeuThrAlaAlaGluMetAla----- 887  
Db 3013 TCTCTCAAAATATCAAGAGAGCTATACAAAGAGAGATGAGAGATGATCAT 3072  
QY 888 -----ProTyrMetIysArgLeuSerArgGlyGlnThrValG1 900  
Db 3073 CAGCAGGTACACGGGCTACTTCTGTGATCTTCAGAAAGCCGCT----- 3118

QY 900 uAspLeuGluValLeuSerAspIle 909  
Db 3119 -GAGTTCATAGAGATCTTTTGGCAT 3145

Search completed: May 14, 2003, 02:39:10  
Job time : 574 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 13, 2003, 23:33:17 ; Search time 314 seconds  
(without alignments)  
4072.431 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 5241  
Sequence: 1 MHTLVVHAWVLTLLGPPRA.....MDSQAQISEALRIILMEAVM 1029

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 783854 seqs, 621352466 residues  
Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n model -DEY=xjh  
-O=/cgn2\_1/USPRO.spool/US09929769/runat\_07052003\_161926\_8130/app.query.fasta.1.1223  
-DB=Published.Applications\_NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62  
-TRANS=Numan40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MTLEN=0  
-MAXLEN=200000000 -USER=US09929769.ecgn.1.1.189.runat\_07052003\_161926\_8130  
-NCPU=6 -ICPU=3 -NO\_XMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA: \*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq: \*  
2: /cgn2\_6/ptodata/2/pubpna/PCIT\_NEW\_PUB.seq: \*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq: \*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq: \*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
6: /cgn2\_6/ptodata/2/pubpna/PCITUS\_PUBCOMB.seq: \*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq: \*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq: \*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq: \*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq: \*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq: \*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq: \*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq: \*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5241	100.0	3437	9	US-10-033-245-21
2	5241	100.0	3437	9	US-10-033-223-21
3	5241	100.0	3437	9	US-10-033-167-21
4	5241	100.0	3437	9	US-10-033-244-21
					Sequence 21, Appl
					Sequence 21, Appl
					Sequence 21, Appl
					Sequence 21, Appl

5	5241	100.0	3437	9	US-10-033-435-21	Sequence 21, Appl
6	5241	100.0	3437	9	US-10-032-990-21	Sequence 21, Appl
7	5241	100.0	3437	9	US-09-929-769-3	Sequence 3, Appl
8	5241	100.0	3437	9	US-10-032-996-21	Sequence 21, Appl
9	5241	100.0	3437	9	US-10-033-396-21	Sequence 21, Appl
10	5241	100.0	3437	12	US-10-033-246-21	Sequence 21, Appl
11	5241	100.0	3437	12	US-10-033-301-21	Sequence 21, Appl
12	5241	100.0	3437	12	US-10-033-326-21	Sequence 21, Appl
13	4211	80.3	2824	10	US-09-822-849A-276	Sequence 276, App
14	2055	39.2	1559	9	US-09-984-271-56	Sequence 56, Appl
15	924	17.6	639	9	US-10-040-739-1131	Sequence 1131, Ap
16	708	13.5	406	9	US-09-918-995-2971	Sequence 2971, Ap
17	592	11.3	484	9	US-09-764-891-1468	Sequence 1468, Ap
18	570	10.9	443	10	US-09-983-965-4270	Sequence 4270, Ap
19	565	10.8	423	10	US-09-983-965-3822	Sequence 3822, Ap
20	374.5	7.1	6413	9	US-10-108-605-26	Sequence 26, Appl
21	317	6.0	436	10	US-09-783-590-9285	Sequence 9285, Ap
22	163.5	3.1	14800	10	US-09-954-456-1601	Sequence 1601, Ap
23	157.5	3.0	7419	10	US-09-815-242-4009	Sequence 4009, Ap
24	152.5	2.9	4080	10	US-09-962-436-342	Sequence 342, App
25	151.5	2.9	3654	9	US-09-938-842A-1225	Sequence 1225, Ap
26	151.5	2.9	7015	9	US-10-132-382-5	Sequence 5, Appl
27	151.5	2.9	7093	9	US-10-132-382-1	Sequence 1, Appl
28	151.5	2.9	7229	9	US-10-132-382-7	Sequence 7, Appl
29	151.5	2.9	7307	9	US-10-132-382-3	Sequence 3, Appl
30	147	2.8	4031	9	US-10-085-108-1	Sequence 1, Appl
31	147	2.8	4121	9	US-09-954-531-131	Sequence 131, App
32	147	2.8	4121	9	US-09-954-531-352	Sequence 352, App
33	145	2.8	4225	9	US-10-085-108-9	Sequence 9, Appl
34	145	2.8	4265	10	US-09-899-651-1	Sequence 1, Appl
35	142.5	2.7	5336	10	US-09-728-952-70	Sequence 70, Appl
36	141	2.7	3774	9	US-10-109-324-1	Sequence 1, Appl
37	139	2.7	15231	10	US-09-917-800A-1505	Sequence 1505, Ap
38	138.5	2.6	35100	10	US-09-782-378A-26	Sequence 26, Appl
39	138	2.6	7995	10	US-09-864-864-333	Sequence 333, App
40	137	2.6	4635	10	US-09-927-112-1	Sequence 1, Appl
41	136	2.6	5023	10	US-09-866-562-14	Sequence 14, Appl
42	135	2.6	8078	10	US-09-791-942-3	Sequence 3, Appl
43	134	2.6	3773	10	US-09-925-302-47	Sequence 47, Appl
44	134	2.6	5466	10	US-09-728-952-69	Sequence 69, Appl
45	134	2.6	7819	10	US-09-964-824A-239	Sequence 239, App

ALIGNMENTS

RESULT 1  
US-10-033-245-21  
; Sequence 21, Application US/10033245  
; Patent No. US20020160392A1  
; GENERAL INFORMATION:  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Godard, Audrey  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2930R1C7  
; CURRENT APPLICATION NUMBER: US/10/033,245  
; CURRENT FILING DATE: 2001-12-27  
; PRIOR APPLICATION NUMBER: 60/095,325  
; PRIOR FILING DATE: 1998-08-04  
; PRIOR APPLICATION NUMBER: 60/112,851  
; PRIOR FILING DATE: 1998-12-16  
; PRIOR APPLICATION NUMBER: 60/113,145

;; PRIOR FILING DATE: 1998-12-16  
;; PRIOR APPLICATION NUMBER: 60/113,511  
;; PRIOR FILING DATE: 1998-12-22  
;; PRIOR APPLICATION NUMBER: 60/115,558  
;; PRIOR FILING DATE: 1999-01-12  
;; PRIOR APPLICATION NUMBER: 60/115,565  
;; PRIOR FILING DATE: 1999-01-12  
;; PRIOR APPLICATION NUMBER: 60/115,733  
;; PRIOR FILING DATE: 1999-01-12  
;; PRIOR APPLICATION NUMBER: 60/119,341  
;; PRIOR FILING DATE: 1999-02-09  
;; PRIOR APPLICATION NUMBER: 60/119,537  
;; PRIOR FILING DATE: 1999-02-10  
;; PRIOR APPLICATION NUMBER: 60/119,965  
;; PRIOR FILING DATE: 1999-02-12  
;; PRIOR APPLICATION NUMBER: 60/162,506  
;; PRIOR FILING DATE: 1999-10-29  
;; PRIOR APPLICATION NUMBER: 60/170,262  
;; PRIOR FILING DATE: 1999-12-09  
;; PRIOR APPLICATION NUMBER: 60/187,202  
;; PRIOR FILING DATE: 2000-03-03  
;; PRIOR APPLICATION NUMBER: PCT/US99/12252  
;; PRIOR FILING DATE: 1999-06-02  
;; PRIOR APPLICATION NUMBER: PCT/US99/28634  
;; PRIOR FILING DATE: 1999-12-01  
;; PRIOR APPLICATION NUMBER: PCT/US99/28551  
;; PRIOR FILING DATE: 1999-12-02  
;; PRIOR APPLICATION NUMBER: PCT/US00/03565  
;; PRIOR FILING DATE: 2000-02-11  
;; PRIOR APPLICATION NUMBER: PCT/US00/04414  
;; PRIOR FILING DATE: 2000-02-22  
;; PRIOR APPLICATION NUMBER: PCT/US00/05841  
;; PRIOR FILING DATE: 2000-03-02  
;; PRIOR APPLICATION NUMBER: PCT/US00/08439  
;; PRIOR FILING DATE: 2000-03-30  
;; PRIOR APPLICATION NUMBER: PCT/US00/14941  
;; PRIOR FILING DATE: 2000-05-30  
;; PRIOR APPLICATION NUMBER: PCT/US00/15264  
;; PRIOR FILING DATE: 2000-06-02  
;; PRIOR APPLICATION NUMBER: PCT/US00/32678  
;; PRIOR FILING DATE: 2000-12-01  
;; NUMBER OF SEQ ID NOS: 38  
;; SEQ ID NO 21  
;; LENGTH: 3437  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-033-245-21  
  
Alignment Scores:  
Pred. No.: 0 Length: 3437  
Score: 5241.00 Matches: 1029  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-09-929-769-7 (1-1029) x US-10-033-245-21 (1-3437)  
QY 1 MethisllleuValIhIsAlaMetValIleuLeuThrLeuGlyProProArgAla 20  
DB 46 ATGCACATCTCTCGTGGCCATGCGATGATCTGTCGACGCGGCGCGCGAGCC 105  
QY 21 AspAspSerGluDheGlnAlaLeuAspIleTrpPheProGluGlyProLeuPro 40  
DB 106 GACGACAGCAGGTTCCAGCGGCTGCTGACATCTGTTCCGAGAGAGCCACTGCC 165  
QY 41 ThrAlaPheLeuValAspTrpSerGluGluAlaLeuLeuProAspTrpPheLeu 60  
DB 166 ACCGCTCTCTGCTGGCAGCATCGAGAGGCGCTGCTTCTCTGACGCTGAACTG 225  
QY 61 ArgMetIleArgSerGluValIleuArgLeuValAspAlaLeuGlnAspLeuGluPro 80  
DB 226 CGCATATATCGTTCTGAGGTGCTCCGCTGCTGAGACCGCCCTGCGAGGACTGAGACCG 285

QY 81 GlnGlnLeuLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerIysLeu 100  
DB 286 CAGAGAGCTGCTGCTGTCGTCAGAGTGTGGATCCCGCTGTCACATGAGAGAACTC 345  
QY 101 LeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGlnAsnIleMet 120  
DB 346 CTCACAGTTCCTGGACCGAGGAGTGGCCACGACCCCGACTGTGGAGAACTCATG 405  
QY 121 AspIysAsnTrpMetAlaHisLeuValGlnIleHisGlnArgGlyAlaSerGlyIle 140  
DB 406 GACAGAGATTACATGGCCACCTGCTGAGAGTCCAGATAGCCGGCGCTCCGAGGC 465  
QY 141 GlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgArgAspSerThrGlnAla 160  
DB 466 CAGACTTCCACTCTTGTGTCACAGCTCCCTGCGCGCCGCGAGACACAGAGAGCA 525  
QY 161 ProIysProIysSerSerProGluGlnProIleGlyIleArgIleArgValGlyThr 180  
DB 526 CCCAAACCAAGAGACAGCCGAGAGACCCCATAGGCCAGGCGGATTCGGGTGGGACC 585  
QY 181 GlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200  
DB 586 CAGCTCGGGGTGGCGGCGCTGAGAGAGACTGGCTGCAATGTTCTCCAGATTTCGCG 645  
QY 201 LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220  
DB 646 CTCAGCCCGGAGCCCTCGGTGGAGAGCTCCAGTCCCGCCGCGGCTGCGCTGCGAG 705  
QY 221 GlnAlaLeuGlyGlnGlnLeuAlaArgValValGlnIleSerProGluValProGlyIle 240  
DB 706 CAGGCGCTGGGCGCAGAGAGCTGGCCGCTGCTGTCAGGAGCCCGAGAGTCCGCGCATC 765  
QY 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyAlaLeu 260  
DB 766 ACGGTGCTGTCTGTCAGGCGCTGCGACCTGCTGACCTGCCACACAGCGGCGCTG 825  
QY 261 ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuCysGln 280  
DB 826 GTGATGTCCATGACACCTGACCTTCCTGCGCTGCGCGCTGCGCGCTGCGAG 885  
QY 281 TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValLeuLeu 300  
DB 886 TACCAAGCGCTGTGTGTCACAGAGACCGGCTTCTCTGCTTCTCGAAGGTCTCTG 945  
QY 301 GlnMetLeuGlnTrpLeuAspSerProGlyValGlnGlyIleProLeuArgAlaGlnLeu 320  
DB 946 CAGATGCTGCAGTGGCTGAGACAGCCCTGGCTGGAGGCGGCGCGGCGCACAGCTC 1005  
QY 321 ArgMetLeuAspSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyIleLeu 340  
DB 1006 AGGATGCTTCCAGACCGCGGCTCAGCGGCGGAGGCTCATGATGTGCGAGGGGCTC 1065  
QY 341 LeuArgLeuAlaGlnAlaLeuAlaPheArgIleAspLeuGlnValValSerSerThrVal 360  
DB 1066 CTGCGCTGGCGCAGGCGCTGCGCTTCTGTCAGAGCTGAGAGGTGCTCAGCTCC 1125  
QY 361 ArgAlaValIleAlaThrLeuArgSerGlyGlnCysSerValGluProAspLeuIle 380  
DB 1126 CGTGGCGTCAATCGCACCTGAGGTGTGGGAGAGTGCAGCGGAGCGGACTGATC 1185  
QY 381 SerIysValLeuGlnGlyLeuIleGlnValArgSerProHisLeuGlnGluLeuLeuThr 400  
DB 1186 AGCAAAATCTCCAGGCGGCTGATGAGTGAAGTCCCGCCACTGAGGAGGAGTGTACT 1245  
QY 401 AlaPheSerSerAlaThrAlaAspAlaAspLeuProPheProAlaCysIysProValVal 420  
DB 1246 GCATCTTCTCTGCTCAGTGGAGTGTGCTCCCGCTTCCAGCTGTAAAGCCGTTGTG 1305  
QY 421 ValValSerSerLeuLeuLeuGlnGlnGluGluProLeuAlaGlyIleProGlyAla 440  
DB 1306 GTGTGAGCTCCCTGCTGCTGCTGAGAGAGAGAGACCCCTGCTGGGGGAGAGCGGGTGC 1365

QY 441 AspglySerLeuAlaValArgLeuGlyProSerSerGlyLeuLeuValAsprr 460  
Db 1366 GAGGTGGACACCTGGAGCGCGTGGCGGGCCCTGTCAGGGCTCTCTATGTGACAG 1425  
QY 461 LeuGluMetLeuAspProGluValValSerSerCysProAspLeuGluLeuArgLeu 480  
Db 1426 CTGGAAAGCTGGACCCGAGGTGGTCAGACAGCTCCCGACCTCCAGCTCAGCTGCTC 1485  
QY 481 PheSerArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 500  
Db 1486 TTCTCCCGAG 1545  
QY 501 ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeu 520  
Db 1546 ACCCTCTTCAGCATCATTCAGCTGGCGCCACACAGCCAGAGTGCATCGAGTCTGCTG 1605  
QY 521 GlySerArgGluGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys 540  
Db 1606 GGCAAGACCGGGAAACAGAGGTTCAGACCCCTCTGCTGAGCTCTCTGCGGCTGCTC 1665  
QY 541 IleHisValProArgIleTropGlnGlyArgAspGlnArgTrpProGlnHisArgArgGlu 560  
Db 1666 ATCCATGTCTCTGCATGTGGCAGGGGGGGGGAGCCAGCCACCCAGAGAGCGGGAG 1725  
QY 561 GluLeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGluLeuIleLeuAla 580  
Db 1726 GAGCTGTGTGGTGGGCTGAGAGGGCCCGAGCTCATGAGCTGTGGAGTGTATCTCGGCC 1785  
QY 581 GluAlaGluTrpThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAlaArg 600  
Db 1786 GAGGGGAGAGCGGAG 1845  
QY 601 LeuProLeuLeuLeuSerCysCysGlyAspAspGlnSerValArgLeuValThrGlu 620  
Db 1846 CTGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1905  
QY 621 HisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArgAsp 640  
Db 1906 CACCTGTGAGCTGCATCAGAGAGTGGGAGAGACAGCTGTGGAGAGGGCTGCCGAGAC 1965  
QY 641 LeuLeuLeuGlnLeuTrpLeuGlnArgProGluLeuArgValProValProGluValLeu 660  
Db 1966 CTCTCTGAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2025  
QY 661 LeuHisSerGlnGlyAlaAlaSerSerSerValCysLeuLeuAspGlyLeuLeuHisArg 680  
Db 2026 CTGCACAGCAGAGGGGCTGCCAGCAGCAGCAGCTGCAAGCTGAGAGACTCATCCAGCC 2085  
QY 681 PheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGluAsnArgGlyAlaAsp 700  
Db 2086 TTCAATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2145  
QY 701 AlaSerMetAlaCysArgGlyLeuAlaValAlaHisProLeuLeuLeuArgHisLeu 720  
Db 2146 GCCAGCATGCGCTGCGGAGAGCTGGGGTGGCGCAGCCGCTGCTGCTGCTGCTGCTG 2205  
QY 721 ProMetIleAlaAlaLeuLeuHisGlyArgTrpHisLeuAspPheGlnGluPheArgGln 740  
Db 2206 CCCATGATCGGGCGCTCTGTCAGCGCGCCAGCCAGCTCACTTCAGAGAGTCCGGCAG 2265  
QY 741 GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuGlnProHis 760  
Db 2266 CAGAACCCACCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2325  
QY 761 ValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgLeu 780  
Db 2326 GTGTCCCGACGAGACACAGGGGGGCTGTGGAGCTGCTGCTGCTGCTGCTGCTGCTG 2385  
QY 781 LeuLeuAsnTrpArgGlySerSerArgHisLeuAlaAlaPheIleAsnLysPheValGln 800  
Db 2386 CTGCTGATTCACAGAACTCCCTCCCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 2445  
QY 801 PheIleHisLysTrpIleThrTyraAsnAlaProAlaAlaIleSerPheLeuGlnLysHis 820

Db 2446 TTCATCCATAAGTACATTACATACATGCCCCAGCAGCATCTCTCTCTGACAGAGCAC 2505  
QY 821 AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu 840  
Db 2506 GCCGACCCGCTCCAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2565  
QY 841 LeuAlaGlyLeuSerLeuProSerArgAspAspArgGlyThrAspArgGlyLeuAspGlu 860  
Db 2566 CTTCGAGGGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2625  
QY 861 GlyGluGluGluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro 880  
Db 2626 GCGGAGAGAGAGAGCTGACAGCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2685  
QY 881 LeuThrAlaAlaGluMetAlaProTrpMetLysArgLeuSerArgGlyGlnThrValGlu 900  
Db 2686 CTGACCGGGCGGAGATGAGCCCTTACAGAAACGGCTTCCCGGGCCAAACGGTGGAG 2745  
QY 901 AspLeuLeuGluValLeuSerAspIleAspGluMetSerArgArgProGluIleLeu 920  
Db 2746 GATCTGCTGAGAGTCTGAGTACATAGACAGATGTCCCGCGAGAGACCCAGATCTCTG 2805  
QY 921 SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGluGluCysCysArgAsn 940  
Db 2806 ACCTTCTCTGACCAACCTGCAGCGGCTGATGAGCTCGGCGAGAGTGTGGCCGCAAC 2865  
QY 941 LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaPhe 960  
Db 2866 CTGCGCTTCAAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2925  
QY 961 LeuProThrPheMetTrpCysLeuGlnSerGlnAspPheGluValValGlnThrAlaLeu 980  
Db 2926 CTGCCACCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2985  
QY 981 ArgAsnLeuProGluTrpAlaLeuLeuLeuCysGlnGlnHisAlaAlaValLeuLeuHisArg 1000  
Db 2986 CGGAACCTCTCGAGTACAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3045  
QY 1001 AlaPheLeuValGlyMetTrpGlyGlnMetAspProSerAlaGlnIleSerGlnAlaLeu 1020  
Db 3046 GCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3105  
QY 1021 ArgIleLeuHisMetGlnAlaValMet 1029  
Db 3106 AGATCTCTCATATGAGGCGCTGATG 3132

RESULT 2  
US-10-033-223-21  
Sequence 21, Application US/10033223  
Patent No. US20020164641  
GENERAL INFORMATION:  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Olang  
APPLICANT: Goddard, Audrey  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2930R1C9  
CURRENT APPLICATION NUMBER: US/10/033,223  
CURRENT FILING DATE: 2001-12-27  
PRIOR APPLICATION NUMBER: 60/095,325  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/112,851



D	b	1306	GAGGTGACGTCCTCGTGGTGCCAGAGAGAGAACCCCTTGCTGGGGAACCCGGGTGCC	1305
O	y	441	ASpGIySerLeuGlAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrp	460
D	b	1366	GAGGTTGGCAGCTGCAGAGGCCGTGGGCTGGGGCCCCCTGTCAGGCCCTCTGATGGACTGG	1425
O	y	461	LeuGlueLettLeuAspProGluValValSerSerCysProAspLeuGlnLeuArgLeuLeu	480
D	b	1426	CTGGAAATGGTGAACCCCGAGGGTGGTCACAGCTGGCCCCGACCTCACTCAAGCTGCTC	1485
O	y	481	PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProTyrlleuLeu	500
D	b	1486	TTCCTCCGGAGAGAAGGGCAAAGSTCAAGGCCAAGTBCCTGCTTCCTGCTCACTCTTG	1545
O	y	501	ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeuLeu	520
D	b	1546	ACCCTCTTCACGCAATCAGTCCAGTCCAGTCCCACACTGCACACATGCATCCGAGTCTGTG	1605
O	y	521	GlyLysSerArgGluGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys	540
D	b	1606	GGCAAGAGCCGGGAACAGAGATTGCACCCCTGCTCTGAGACTTCTCTGGGCCCTGC	1665
O	y	541	IleHisValProArgIleTrpGlnGlyArgAspGlnArgTrpProGlnLysArgArgGlu	560
D	b	1666	ATCCATGTTCCTGCTGCATCTGGCAGGGGGCCGGACACGCAACCCCAAGACGGCGGAG	1725
O	y	561	GluLeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGlnLeuIleLeuAla	580
D	b	1726	GAGCTGTGCTGGGGGTCAGAGGCCCGGAGATCATCAGCTGTGTGAGTGTACTTGCC	1785
O	y	581	GluAlaGluThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAlaArg	600
D	b	1786	GAGCGGAGACGGGAGCCAGGACGGGGACACAGCCGCTGACGCTCATCCAGGCCGG	1845
O	y	601	LeuProLeuLeuLeuSerCysCysGlyAspAspGlnSerValArgLysValThrGlu	620
D	b	1846	CTGCGCCCTGCTGCTCACTCTCTGCTGGGAGCATGAGATGTGACGAAAGTGAACGAG	1905
O	y	621	HisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArgAsp	640
D	b	1906	CACCTGCAGAGCTGCATCCAGACAGTGGGGAGACAGCGTCTGGAAAGGGGCTGCCGAGAC	1965
O	y	641	LeuLeuLeuGlnLeuTrpLeuGlnArgProGluLeuArgValProValProGluValLeu	660
D	b	1966	CTTCTCTGCAGCTCTACCTACACAGCCGAGCACTCGGGGTCCCGTGCCTGAGGTCTTA	2025
O	y	661	LeuHisSerGlnGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuLeuHisArg	680
D	b	2026	CTGCACACGGAAGGGCTGCCACAGACAGCGTGTGCAACTGCAGCGACTCTCCACCGC	2085
O	y	681	PheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAlaAsp	700
D	b	2086	TTCATCACGGCTCCTGGCGACACACAGACTCCCGGGGCTTGAGAACGAGAGGGCGGAT	2145
O	y	701	AlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuLeuArgHisLeu	720
D	b	2146	GCCAGCATGGCTCTGCCGGAAGCTGGCGGTGGCCACACCCTGCTGCTCAGCACCTGG	2205
O	y	721	ProMetIleAlaAlaLeuLeuHisGlyArgTrpHisLeuAsnPhgGlnGlnPheArgGln	740
D	b	2206	CCCATGATCGGGGGCTCCGACAGGCCGAGACCCATCACTCACTTCAGAGATTCGGGAG	2265
O	y	741	GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnIleuLeuGlnProHis	760
D	b	2266	CAGAACCACTGAAGCTCTCTCCGACAGTGTGGGCTGCTGAGACTGCTGAGCGCGAC	2325
O	y	761	ValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgLeu	780
D	b	2326	GTTGTTCCGACAGCGACCCAGAGGGGGCGCTGTGGACTGTGCTTCGTTCATCCGCTGT	2385
O	y	781	LeuLeuAsnTrpArgLysSerSerArgHisLeuAlaAlaPheIleasnLysPheValGln	800
D	b	2386	CTGTGTGAATTACAGGAAGTCTCCCGCACTGTGGCTGCTTATACAAAGTTTGTCGAG	2445

Oy	801	PheIleHstYstrYrTrieThrrYrAsnAlaProAlaAlaIleSerPheLeuLnyHis	820
Db	2446	TTCATTCATAAGTACATTACTCTCAATGCCCCAGCAGCCATCTCTTCTGTGAGAAGAC	2505
Oy	821	ALAASPProLeuHisAspPLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu	840
Db	2506	GCCGACCCTGCCACGACCTGTCTTGACAACAGTAGACTGGTGAATGCTGAATCCCTC	2565
Oy	841	LeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArgGlyLeuAspGluGln	860
Db	2566	CTTGACGAGGCTCACCTCGCCACAGCAGGAGCAGACGACGACCGAGGCTGTGACGAAGAG	2622
Oy	861	GlyGluGluGlnUserSerAlaGlySerLeuProLeuValSerAlaSerLeuPheThrPro	880
Db	2626	GCCGAGGAGGAGACTCATGGCGGCTCTTGTCCCCTGTGTACGCTCTTCTTTTCACCCCT	2683
Oy	881	LeuThrAlaAlaGluMetAlaProTyMetLysArgLeuSerArgGlyGlnThrValGlu	900
Db	2686	CTGACCGGGGCGGAGATGGCCCTCATATGAACGGCTTCCCGGGGCCAAGCGGTGGAG	2742
Oy	901	AspLeuLeuGluValLeuSerAspLileaspGluMetSerArgArgArgProGluLeu	920
Db	2746	GATCTGCTGGAGGTYCTTGAGTGCATGTAGACGAGATGTCCCGGGGAGACCCGAGATCTCG	2805
Oy	921	SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGluGluGlyCysArgAsn	940
Db	2806	AGCTTCTTCTTGACCAACCTGACACGGGCTGTATGAGTCTCGGCCGAGGAGTGTGGCGAAC	2865
Oy	941	LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaPhe	960
Db	2866	CTCGGCTTCACACTGGGCGCTGGCTCCATGCAGAACAGCCCCAGCATTCACGCGGTTTTG	2922
Oy	961	LeuProthrPheMetTrpCysLeuGlySerGlnAspPheGluValValGlnThrAlaLeu	980
Db	2926	CTGGCCACGTTTCATGACTGCTGTGGCAGCAGAGACTTGTAGGTGTGTGCAGACGGGCTTC	2983
Oy	981	ArgAsnLeuProGluTrpAlaLeuLeuLeuCysGlnGluHisAlaAlaValLeuLeuHisArg	1000
Db	2986	CGGAACTCGCTGTAGTACGCTCTCTCTGTCCMAAGACGACCGGGCTGTGCTCCACCGG	3042
Oy	1001	AlaPheLeuValGlyMetTrpGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeu	1020
Db	3046	GCCTTCCGTGGGGCAGTAGTACGGCCAGATGAGACCCGACGCGAGATCTCCGAGGCGCTG	3103
Oy	1021	ArgIleLeuHisMetGluAlaValMet	1029
Db	3106	AGGATCTGCATATGAGAGCGCGTGATG	3132
RESULT 3			
US-10-033-167-21			
Sequence 21, Application US/10033167			
Publication No. US20020182618A1			
GENERAL INFORMATION:			
APPLICANT: Botstein, David			
APPLICANT: Desnoyers, Luc			
APPLICANT: Ferrara, Napoleone			
APPLICANT: Fong, Sherman			
APPLICANT: Gao, Wei-Qiang			
APPLICANT: Goddard, Audrey			
APPLICANT: Gurney, Austin L.			
APPLICANT: Pan, James			
APPLICANT: Roy, Margaret Ann			
APPLICANT: Stewart, Timothy A.			
APPLICANT: Tumas, Daniel			
APPLICANT: Watanabe, Collin K.			
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic			
FILE REFERENCE: P2930R1C10			
CURRENT APPLICATION NUMBER: US/10/033,167			
PRIOR FILING DATE: 2001-12-27			
PRIOR APPLICATION NUMBER: 60/095,325			

```

: PRIOR FILING DATE: 1998-08-04
: PRIOR APPLICATION NUMBER: 60/112,851
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,145
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,511
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: 60/115,558
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/115,565
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/115,733
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/119,341
: PRIOR FILING DATE: 1999-02-09
: PRIOR APPLICATION NUMBER: 60/119,537
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: 60/119,965
: PRIOR FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: 60/162,506
: PRIOR FILING DATE: 1999-10-29
: PRIOR APPLICATION NUMBER: 60/170,262
: PRIOR FILING DATE: 1999-12-09
: PRIOR APPLICATION NUMBER: 60/187,202
: PRIOR FILING DATE: 2000-03-03
: PRIOR APPLICATION NUMBER: PCT/US99/12252
: PRIOR FILING DATE: 1999-06-02
: PRIOR APPLICATION NUMBER: PCT/US99/28634
: PRIOR FILING DATE: 1999-12-01
: PRIOR APPLICATION NUMBER: PCT/US99/28551
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US00/03565
: PRIOR FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: PCT/US00/05841
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: PCT/US00/08439
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: PCT/US00/14941
: PRIOR FILING DATE: 2000-05-30
: PRIOR APPLICATION NUMBER: PCT/US00/15264
: PRIOR FILING DATE: 2000-06-02
: PRIOR APPLICATION NUMBER: PCT/US00/32678
: PRIOR FILING DATE: 2000-12-01
: NUMBER OF SEQ ID NOS: 38
: SEQ ID NO 21
: LENGTH: 3437
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-033-167-21

Alignment Scores:
Pred. No.: 0
Score: 5241.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 9
Length: 3437
Matches: 1029
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-929-769-7 (1-1029) x US-10-033-167-21 (1-3437)
QY 1 MethisthaleuValHlHsAlaMetValHleuLeuThrlLeuGlyProProAla 20
Db 46 ATGACACTCTCGTGGTCCATGCGCATGATCTGCTGAGCGCTGGCGCCCTCGAGCC 105
QY 21 AspAspserGluPheGlnAlaLeuAspLleThrpPheProGluGlyProLeuPro 40
Db 106 GACACACGCGAGTTCAGGCGCTCTGACATCTGTTCCGGAGGAGAACCCACTGCC 165
QY 41 ThrAlaPheLeuValAspThrSerGluGlyAlaLeuLeuLeuProAspThrLeuLysLeu 60
Db 166 ACCGCTTCTCGTGGACACATCGAGAGCGCTGCTGCTTCTGACTGCTGACTG 225

```

```

QY 61 ArgMetIleArgSerGluValLeuArgLeuValAspAlaAlaLeuGlnAspLeuGluPro 80
Db 226 CGCATGATTCGCTGTGAGTCTCTCGCTGTGTGAGCGCGCCGCGAGACTGAGACCG 285
QY 81 GlnGlnLeuLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerLysLeu 100
Db 286 CAGCAGCTGCTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 345
QY 101 LeuGlnPheLeuAspGlnAlaValAlaHlsAspProGlnThrlLeuGlnGlnAlaMet 120
Db 346 CTCAGTTCCTGAGCAGCAGCTGCGCCAGCAGCCAGCAGCTGAGCAGACATCATG 405
QY 121 AspLysAsnTyrMetAlaHlsLeuValGluValGlnHlsGluArgGlyAlaSerGlyGly 140
Db 406 GACAAGATTCATGAGCGCCAGCTGTGAGAGTCCAGCATGAGCGCGCGCTCGGAGGC 465
QY 141 GlnThrPheHlsSerLeuThrlAlaSerLeuProProAlaArgAspSerThrGlnAla 160
Db 466 CAGACTTCCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
QY 161 ProLysProLysSerSerProGluGlnProIleGlyGlnArgValGlyThr 180
Db 526 CCAAAACCAAGAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 585
QY 181 GlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200
Db 586 CAGCTCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 645
QY 201 LeuSerProAspProAlaArgThrGlnSerSerSerProAlaProValAlaLeuGln 220
Db 646 CTCAGCCGCGACCTCGTGGTGGAGAGCTCCAGTCCGCGCGCGCGCGCTGCGCTGAG 705
QY 221 GlnAlaLeuGlyGlnGlnLeuAlaArgValGlnGlnGlnSerProGluValProGlyIle 240
Db 706 CAGGCGCTGGCGCAGAGAGTGGCGCGCGCTGCTGAGGAGCGCGCGCGCGCGCGCTG 765
QY 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerProHlsGlyAlaLeu 260
Db 766 ACGGTGGTGTCTGCGAGCGCTGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 825
QY 261 ValMetSerMetHlsArgSerHlsPheLeuAlaLysProLeuLeuArgGlnLysGln 280
Db 826 GTGATGTCATGACCGCTAGCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 885
QY 281 TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValLeuLeu 300
Db 886 TACCAGCGCTGTGTGCGCAGAGACACCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 945
QY 301 GlnMetLeuGlnThrPheAspSerProGlyAlaGlnGlyGlyProLeuArgAlaGlnLeu 320
Db 946 CAGATGCTGCGAGTGGTGGAGAGCGCTGCGTGGAGGCGCGCGCGCGCGCGCGCTG 1005
QY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeu 340
Db 1006 AGGATGCTTCCAGCAGCAGCTGAGCGCGCGCGCGCGCTGAGTGTGCGAGGCGCGCTC 1065
QY 341 LeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGluValSerSerThrVal 360
Db 1066 CTGCGCGTGGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
QY 361 ArgAlaValIleAlaThrLeuArgSerGlyGlnGlnGlnSerValGluProAspLeuIle 380
Db 1126 CGTCCGCTGATCCAGCAGCAGCTGAGTGTGGGAGAGAGCGAGCGGAGCGGAGCTATC 1185
QY 381 SerLysValLeuGlnGlnGlyLeuIleGluValArgSerProHlsLeuGlnGlnLeuThr 400
Db 1186 AGCAAGTCTTCCAGGCGCTGATGAGGTAGGTCCCGCAGCTGAGAGAGAGCTCTACT 1245
QY 401 AlaPhePheSerAlaThrAlaAspAlaLaserProPheProAlaCysLysProValVal 420
Db 1246 GCATTCTTCTGCGACATGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305

```



OY	421	VAlVAlSerSerLeuLeuLeuGlnGlnGluGluProLeuAlaGlyGlyLysProGluA	440
Dd	1306	GtGGTGAAGCTCCCTGGCTGTGCACGAGAGAGAACCCCTTGCTGGGGGAGAAGCGGGTTCG	1365
OY	441	ASpAlglySerLeuGlnAlaValArGLeuGlyProSerSerGlyLeuLeuValAsPrp	460
Dd	1366	GAGCGTGCAAGCCTGAGGCCGTCGGGGCTGGGGCCCTGTCAGAGGCTCCTATGACATCG	1425
OY	461	LeuGluMetLeuAspProGluValValSerSerCysProAspLeuGlnLeuArgLeuLeu	480
Dd	1426	CTGGAAATGCTGGACCCCCGAGGTGGTTCACAGCTGCCCGCATCTCACGTCAAGCTCTGC	1489
OY	481	pHeSerArgrArgLysGlyLysGlyGlnAlaGlnValProSerPheArpProTyrtLeuLeu	500
Dd	1486	TTCCTCCGGAGAGAGGCAAAAGTCAAGGCCACAGGTCCCTCTGCTCCGTTCACTCTCTG	1545
OY	501	ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlyIleArGValLeuLeu	520
Dd	1546	ACCCTCTCACAGCATCAGTCCACTGGCCCAACATGCACATGCATGCATGCATGCTGGCTG	1605
OY	521	GlyLysSerArgGluGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys	540
Dd	1606	GGCAAGAGCGCGAACAAGAGTTTCGACCCCTCTGCTCTCTGAGATTCTCTGGCGCTTC	1665
OY	541	IleHisValProArgIleTrpGlnArgAspGlnArgTrpProGlnLysArgArgLys	560
Dd	1666	ATTCATGTTCTCTGCATCTCTGGCAAGGGCGGGACACAGCAACCCCGCAGAAAGCGGGAG	1725
OY	561	GluLeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGluLeuLeuAla	580
Dd	1726	GAGTGTGGTGGCGGGTCCAGAGGCCCGCGAGCTCATAGCTGTGTGAGCTGATCTTGCC	1785
OY	581	GluAlaGluThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAlaArg	600
Dd	1786	GAGCGGAGAGCGCGAGACCAGAGCGGGAGACACAGCCGCTCAGACTCATCCAGACCCGG	1845
OY	601	LeuProLeuLeuLeuSerCysCysGlyAspAspGluSerValArgLysValThrGlu	620
Dd	1846	CTGGCCCTGGCTCACCTCTCTCTGGGGAGAGAAAGTCTCAGAAAGGTGACAGAG	1905
OY	621	HisLeuSerGlyCysIleGlnGlnInTPrgLYAspSerValLeuGlyARgArgCysArgSp	640
Dd	1906	CACCTGTACAGCTCATCCACAGCATGGGGAGACAGCGTGTGGGAAGGGCTGCCGAGAC	1965
OY	641	LeuLeuLeuGlnLeuTyrlLeuGlnArgProGluLeuArgValProValProGluValLeu	660
Dd	1966	CTTCTCTGACGCTCTACACAGCGGCGGAGACTCGGGGTGCCCTGACGAGGTCTCA	2025
OY	661	LeuHisSerGluGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuIleHisArg	680
Dd	2026	CTGCACAGCCAAAGGGCTGCCACACAGCGTCTGCACACTGGACGAGACTCATCCACCGC	2085
OY	681	PheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnLysAnrGlyValAsp	700
Dd	2086	TTCATCACGCTCTTGGCGGACACCGAGACTCCCGGGCTTGAGAGACCGAGGGCGGAT	2145
OY	701	AlaSerMetaLacysArgLysLeuAlaValAlaHisProLeuLeuLeuLeuArgHisLeu	720
Dd	2146	GCCAGCATGGGCTGCCGAAGCGTGGGGTGGCCACCGCTGTGTGTCTCAGGACACTG	2205
OY	721	ProMetIleAlaAlaLeuLeuHisGlyArgTrpThrHisLeuAsnPhcGlnIleuPheArgLn	740
Dd	2206	CCCATTGATCGGGGCTCTCGCACAGGCGCGCACCCCATCACTTCACAGAGATTCCGGCAG	2265
OY	741	GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuGlnProHis	760
Dd	2266	CAGAACCACTGACCTGTCTTCCGACAGTGTGGCTGTCTGTGAGTGTGTGACGCGGAC	2325
OY	761	ValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgLeu	780
Dd	2326	GtGTTCGCGACGAGCACAGAGGGGGCTGTGGGAACTGCTTCTGTCTTATCCGCTG	2385
OY	781	LeuLeuAsnTyrrArgLysSerSerArgHisLeuAlaAlaPheIleAsnLysPheValGln	800

Db	2386	CTGTGGATTACAGGAAGTCTCCGCCATCTGGCTTGCCTTCATTAACAAGTTTGTGAG	2445
QY	801	PhelIleHslsYrYrIleThrYTrAsnAlaProAlaAlaIleSerPheLeuGlnLysHis	820
Db	2446	TTTCATTCATTAAGTATACATTACTACATATGCCCCAGCAGCATCTCTTCTCCACAAAGAC	2505
QY	821	AlaAspProLeuHtsAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu	840
Db	2506	GCCACACCGGCTCCACAGCCCTGTCCTTCGCAACAGTACCTGTATCTGAATATCCCTC	2565
QY	841	LeuAlaGlyLeuSerLeuProSerArgAspAspArgTrhAspArgGlyLeuAspGln	860
Db	2566	CTTCACAGGCTCAGCCCTGCCACAGGAGCAGACAGGACCGACCGAGGCTTGACCAAGAG	2625
QY	861	GlyGlnGlnGlnSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro	880
Db	2626	GCCGAGAGAGAGACTCAGCCGCCGCTCCCTGACCCCTGTCAGCGTCTCCCTGTACCCCT	2685
QY	881	LeuThrAlaAlaGlnMetAlaProTyrMetLysArgLeuSerArgGlyGlnThrValGlu	900
Db	2686	CTGACCCCGGCGCGAGGAGGCCCTTCACATGAACACGGCTTTCCGGGGCCAAACGGTGGAG	2745
QY	901	AspLeuLeuGlnValLeuSerAspIleAspGlnMetSerArgArgArgProGlnLeu	920
Db	2746	GATCTGCTGGAGGTTCTGAGTGCATAGACAGATATCCCGGGCGAGACCGAGATCCTG	2805
QY	921	SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGlnGlnLysCysArgAsn	940
Db	2806	AGCTTCCTTCGACCAACCTGACGGCGGTGATGTAGCTCGCGGAGAGAGTTGCCCGAC	2865
QY	941	LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaPhe	960
Db	2866	CTCCCTTCACAGCCCGGCGCTGCTCCATGCAAGACGCCCAAGCATTTGCAGCGCTTTC	2925
QY	961	LeuProThrPheMetTyrCysLeuGlnGlnSerGlnAspPheGlnValValGlnThrAlaLeu	980
Db	2926	CTGGCCACGTTCACTGATCTCCTGGGCGAGCAGAGATTTGAGTGATGTGCAAGCGGCTTC	2985
QY	981	ArgAsnLeuProGlnTyrAlaAlaLeuLeuCysGlnGlnHisAlaAlaValLeuLeuHtsArg	1000
Db	2986	CGGAACCTGCTGAGTACGCTCCCTGTCGCAAGACACGGCGCTGTCTGTCCACCGG	3045
QY	1001	AlaPheLeuValGlyMetTyrGlnMetAspProSerAlaGlnIleSerGlnAlaLeu	1020
Db	3046	GCCCTCTCTGGTGGGCATGTACGGCCAGATGAGACCCAGCGCGCAGATCTCCGAGGCCCTG	3105
QY	1021	ArgIleLeuHtsMetGlnAlaValMet	1029
Db	3106	AGGATCCTGCATATGAGGCGCGTGAATG	3132
RESULT 4			
US-10-033-244-21			
: Sequence 21, Application US/10033244			
: Publication No. US20020192686A1			
: GENERAL INFORMATION:			
: APPLICANT: Botstein, David			
: APPLICANT: Desnoyers, Inc			
: APPLICANT: Ferrara, Napoleone			
: APPLICANT: Fong, Sherman			
: APPLICANT: Gao, Wei-Qiang			
: APPLICANT: Goddard, Audrey			
: APPLICANT: Gurney, Austin L.			
: APPLICANT: Pan, James			
: APPLICANT: Roy, Margaret Ann			
: APPLICANT: Stewart, Timothy A.			
: APPLICANT: Tumas, Daniel			
: APPLICANT: Watanabe, Colin K.			
: APPLICANT: Wood, William I.			
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic			
: TITLE OF INVENTION: Acids Encoding the Same			
: FILE REFERENCE: P2930R1C2			
: CURRENT APPLICATION NUMBER: US/10/033,244			

```

CURRENT FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: 60/095,325
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/112,851
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,145
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,511
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/115,558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119,341
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119,537
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/119,965
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: 60/162,506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170,262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/187,202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/14941
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
NUMBER OF SEQ ID NOS: 38
SEQ ID NO 21
LENGTH: 3437
TYPE: DNA
ORGANISM: Homo sapiens
US-10-033-244-21

Alignment Scores:
Pred. No.: 0 Length: 3437
Score: 5241.00 Matches: 1029
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-929-769-7 (1-1029) x US-10-033-244-21 (1-3437)
QY 1 MethisIleuValValHisAlaMetValIleLeuLeuThrLeuGlyProPArgAla 20
DB 46 ATGCACATCTCTGTCGTCATGCCATGTCCTGCTGACGCTGGGCCCGCTCGAGCC 105
QY 21 AspAspSerGluPheGlnAlaLeuLeuAspIleTirPheProGluGlyProLeuPro 40
DB 106 GACGACAGCGAGTTCACAGGGCTGTCGACATCTGTTCCGAGAGAGAACCCAGCC 165
QY 41 ThrAlaPheLeuValAspThrSerGluGluAlaLeuLeuProAspThrLeuLysLeu 60

```

```

DB 166 ACCGCTTCTCTGTCGTCACATCGAGAGAGCCCTGCTCTCTGCTGACGTCGAGACTG 225
QY 61 ArgMetIleArgSerGluValIleArgLeuValAspAlaIleLeuGlnAspLeuGluPro 80
DB 226 CGCATGATCCGTTCTGAGGTCCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 285
QY 81 GlnGlnLeuLeuLeuPheValGlnSerPheGlyTleProValSerSerMetSerLysLeu 100
DB 286 CAGCAGCTGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 345
QY 101 LeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGluGlnAsnIleMet 120
DB 346 CTCACGTTCTGACACAGGCGAGCGGCCACGACCCGACCTGTGAGCAGAACATCATG 405
QY 121 AspLysAsnTyrMetAlaHisLeuValGlnGlnHisGlnArgGlyAlaSerGlyGly 140
DB 406 GACACAGATTACATGCCCCCTGTCGAGGTTCACAGTACAGCGCCGCTCCGAGGCG 465
QY 141 GlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgArgAspSerThrGluAla 160
DB 466 CAGACTTCCACTCTGTCACAGCCTCCGTCGCGCCGCGCGAGACACACAGAGGCA 525
QY 161 ProLysProLysSerSerProGluGlnProIleGlyGlnGlyArgTleArgValGlyThr 180
DB 526 CCCAAACCAAGAGACGCCAGAGACCCCATAGGCCAGGCGCGATCGGGTGGGAGCC 585
QY 181 GlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200
DB 586 CAGCTCGGGTGTGGGCCCTGAGAGACGACTGTCGTCGTCGTCGTCGTCGTCGTCGTCG 645
QY 201 LeuSerProAspProArgTirGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220
DB 646 CTCAGCGCGGACCTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 705
QY 221 GlnAlaLeuGlyGlnGlnIleAlaArgValGlnGlySerProGluValProGlyTle 240
DB 706 CAGGCTTGGGCCAGAGAGCTGCGCGCTGCTGTCAGAGGACGCCCGAGGTGCGGGCATC 765
QY 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyAlaLeu 260
DB 766 ACGGTGCTGTCCTGACGAGCGCTGCGCACCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
QY 261 ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuCysGln 280
DB 826 GTATGTCATGACCGTCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885
QY 281 TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuLysValLeuLeu 300
DB 886 TACACGCGCTGTGTCACAGACACCGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
QY 301 GlnMetLeuGlnThrLeuLysSerProGlyValGlnGlyGlyProLeuArgAlaGlnLeu 320
DB 946 CAGATGCTGACGTGTCGACAGACACCGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005
QY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeu 340
DB 1006 AGGATGCTTGCACGACGCGCTGACGCGGCGGACGCTCAGTATGTGCGAGGGGGCTC 1065
QY 341 LeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValValSerSerThrVal 360
DB 1066 CTCGCGCTGGCGAGGCGCTGCGCTTCTGCTGAGGACTGAGAGTGGTCACTCCACCGTC 1125
QY 361 ArgAlaValIleAlaThrLeuArgSerGlyGlnGlnCysSerValGlnProAspLeuLeu 380
DB 1126 CGTGGCGTATGCCACACCTGAGGTCGTGGGAGACAGTGCAGGTGAGCGGACGTCGATC 1185
QY 381 SerLysValLeuGlnGlnLeuIleGluValArgSerProHisLeuGlnGlnLeuLeuThr 400
DB 1186 AGCAAGTCTCTCCAGGGGCTGATGAGGTGAGGTCCCGACCTGAGAGAGCTGCTGACT 1245
QY 401 AlaPhePheSerLanThrLanAspAlaLaserProPheProAlaCysLysProValVal 420

```



```

FILE REFERENCE: P2930R1C5
CURRENT APPLICATION NUMBER: US/10/033.435
CURRENT FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: 60/095.325
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/112.851
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113.145
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113.511
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/115.558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115.565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115.733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119.341
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119.537
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/119.965
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: 60/162.506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170.262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/187.202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/14941
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
NUMBER OF SEQ ID NOS: 38
SEQ ID NO 21
LENGTH: 3437
TYPE: DNA
ORGANISM: Homo sapiens
US-10-033-435-21

Alignment Scores:
Pred. No.: 0
Score: 5241.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 9
Gaps: 0

US-09-929-769-7 (1-1029) x US-10-033-435-21 (1-3437)
QY 1 MetHisIleuValValHisAlaMetValIleleuLeuThrLeuGlyProProArgAla 20
Db 46 ATGCACATCTCTGCGGTCATGCGCATGCTGCTGACGCGTGGCCGCTCGAGCC 105
QY 21 AspAspSerGluPheGlnAlaLeuLeuAspIleTrpPheProGluGluProLeuPro 40
Db 106 GAGCAGACGAGGATTCAGGCGCTGCTGACATCTGTTCCGAGAGGAGAACCCACTGCC 165

```

```

QY 41 ThrAlaPheLeuValAspThrSerGluGluAlaLeuLeuProAspTrpLeuGlyLeu 60
Db 166 ACCGCTTCTCTGGTGGACATCGAGAGGCGGCTGCTGCTTCTGATCGATGAACCTG 225
QY 61 ArgMetIleArgSerGluValLeuArgLeuValAspAlaAlaLeuGlnAspLeuGluPro 80
Db 226 CGCATGTATCCGTTCTGAGGTCCTCCGCTGAGACCCCGCTGACAGACCTGGAGCCG 285
QY 81 GlnIleLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerLeu 100
Db 286 CAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 345
QY 101 LeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGluGlnAspIleMet 120
Db 346 CTCAGTTCCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 405
QY 121 AspLysAsnTrpMetAlaHisLeuValGluValGlnHisGluArgGlyAlaSerGlyGly 140
Db 406 GACAGAGATTACATGATGCCACCTGCTGAGTCCAGATGAGCGCGGCTCCGAGGCG 465
QY 141 GlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgAspSerThrGluAla 160
Db 466 CAGACTTTCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
QY 161 ProLysProLysSerSerProGluGlnProIleGlyGlnGlyArgIleArgValGlyThr 180
Db 526 CCCAAGACCAAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 585
QY 181 GlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200
Db 586 CAGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 645
QY 201 LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220
Db 646 CTCAGCCCGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 705
QY 221 GlnAlaLeuGlyGlnGluLeuAlaArgValAlaGlnGlySerProGluValProGlyIle 240
Db 706 CAGGCTCGGCGCCAGGAGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765
QY 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyAlaLeu 260
Db 766 AGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
QY 261 ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuGlyGln 280
Db 826 GTGATGTCCATGCACGCTAGCAGCTTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885
QY 281 TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValLeuLeu 300
Db 886 TACAGCGCTGTGTGCGCACAGGACACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
QY 301 GlnMetLeuGlnTrpLeuAspSerProGlyValGlnGlyGlyProLeuArgAlaGlnLeu 320
Db 946 CAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005
QY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeu 340
Db 1066 AGATGCTTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1065
QY 341 LeuArgLeuAlaGluAlaLeuAlaPheArgGlnAspLeuGluValAlaSerSerThrVal 360
Db 1066 CTGCGCTGCGCGAGGCGCTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
QY 361 ArgAlaValAlaIleAlaThrLeuArgSerGlyGlnGlnCysSerValGluProAspLeuIle 380
Db 1126 CGTGCCTGATCGCACCTGAGGCTGCGGAGCAGTGCAGCTGAGCGGACCTGATC 1185
QY 381 SerLysValLeuGlnGlyLeuIleGluValAlaArgSerProHisLeuGlnGluLeuThr 400
Db 1186 AGCAAACTCTCCAGAGGCGCTGATGAGGTGAGGTCCCGACCTGAGGAGAGCTGCTGACT 1245

```

QY 401 AlaPheSerAlaThrAlaAspAlaAspProPheProAlaCysLysProValVal 420  
 DB 1246 GCATTCTCTCTGCGACGCGGATCTGCTCCCGTTCCAGCTGTAAGCCCTGTGG 1305  
 QY 421 ValValSerSerLeuLeuLeuGlnGlnGlnGlnProLeuAlaGlyLysProGlyAla 440  
 DB 1306 GTGGTGACCTCCCTGCTCTGCGAGAGAGAGCCCTGGCTGGGGAGAGCCGGGTGG 1365  
 QY 441 AspGlyGlySerLeuGlnAlaValAlaArgLeuGlyProSerSerGlyLeuLeuValAspTrp 460  
 DB 1366 GACGGTGACGCTGGAGAGCGCTGGCGGGCCCTGTCAGCGCTCTCTAGTGAAGCTGG 1425  
 QY 461 LeuGlnMetLeuAspProGlnValValSerSerCysProAspLeuGlnLeuArgLeuLeu 480  
 DB 1426 CTGGAATAGCTGAGCCCGAGGTGGTCAAGACCTGCCGACCTGACGTCAAGCTGCTC 1485  
 QY 481 PheSerArgArgGlyGlyLysGlyGlnAlaGlnValProSerPheArgProTrpLeuLeu 500  
 DB 1486 TTCTCCCGAGAGAAAGGGAAGGTGAGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545  
 QY 501 ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeuLeu 520  
 DB 1546 ACCCTTCTCAGCATGATCCAGCTGGCCACACTGACACAGTGCATCCGAGTCTGCTG 1605  
 QY 521 GlyLysSerArgGlnGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys 540  
 DB 1606 GGCAGAGACCGGGAGACAGAGGTGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1665  
 QY 541 IleHisValProArgIleTrpGlnGlyArgAspGlnArgTrpProGlnLysArgArgGlu 560  
 DB 1666 ATCCATGCTCTCGCATGTGGCAGAGGGGGGAGCAGCCAGCCAGCCAGAGAGGGGGAG 1725  
 QY 561 GlnLeuValLeuArgValGlnGlyProGlnLeuIleSerLeuValGlnLeuLeuAla 580  
 DB 1726 GAGCTGGTGTCTGCGGCTCAGGAGCCCGAGACTCATCAGCTGTGTGAGCTGATCTG 1785  
 QY 581 GlnAlaGlnThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAlaArg 600  
 DB 1786 GAGCGGAGAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1845  
 QY 601 LeuProLeuLeuLeuSerCysCysCysGlyAspAspGlnSerValArgLysValThrGlu 620  
 DB 1846 CTGCCCCCTGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1905  
 QY 621 HisLeuSerGlyCysIleGlnGlnTrpGlnAspSerValLeuGlyArgArgCysArgAsp 640  
 DB 1906 CACCTGTAGGCTGCATTCAGAGTGGGAGACAGCTGTCTGGAGAGCGCTGCCGAGAC 1965  
 QY 641 LeuLeuLeuGlnLeuTrpLeuGlnArgProGlnLeuArgValProValProGlnValLeu 660  
 DB 1966 CTTCCTGTCAGCTCTACCTACAGCGCGGAGCTGGGGGTGCCGTGCTGAGGTCTTA 2025  
 QY 661 LeuHisSerGlnGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuLeuHisArg 680  
 DB 2026 CTGCAACAGCGAAGGGGTGCTCCAGCAGACGCTCTCAACCTGAGGAGTCTATCCACC 2085  
 QY 681 PheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnAlaAspArgGlyAlaAsp 700  
 DB 2086 TTCACTCAGGCTCTTGGCGAGACACAGCACTCCCGGGCTTGGAGAGACGAGGGCGGAT 2145  
 QY 701 AlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuLeuArgHisLeu 720  
 DB 2146 GCCACATAGGCTCTGCGGAGGCTGGGCGGAGCCCGCTGCTGCTGCTGAGCAGCTG 2205  
 QY 721 ProMetIleAlaAlaLeuLeuHisGlyArgThrHisLeuAspPheGlnIlePheArgGln 740  
 DB 2206 CCCATGATGCGGGCTCTGCGAGCGCGCAGCCCACTCACTTCCAGAGAGTTCGGGCGAG 2265  
 QY 741 GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuLeuInProHis 760  
 DB 2266 CAGAACCACTGTAGCTGCTCTGCAAGCTGTGGGGCTGTGGAGCTGTGTGACAGCGGAC 2325  
 QY 761 ValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgLeu 780

DB 2326 GTGTTCCGAGGAGAGACAGGAGGGGGGTGTGGAGCTGCTCTGCTCTCATCCGCGCTG 2385  
 QY 781 LeuLeuAsnTrpArgLysSerSerArgHisLeuAlaAlaPheIleAsnLysPheValGln 800  
 DB 2386 CTGCTGAATTTACAGGAAGCTCTCCGCCATCTGCGCTGCTCATCAACAAGTTTGTGAG 2445  
 QY 801 PheIleHisLysTrpIleThrTyraAsnAlaProAlaAlaIleSerPheLeuGlnLysHis 820  
 DB 2446 TTCACTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2505  
 QY 821 AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu 840  
 DB 2506 GCCGACCGCTCCAGAGACTGATCTTCGACAAACAGAGACTGTGTATGCTGAATCCCTTC 2565  
 QY 841 LeuAlaGlyLeuSerLeuProSerArgAspAspArgTrpHisAspArgLysLeuAspGlnGlu 860  
 DB 2566 CTGGCAGGGCTCAGCTGCTGCCAGAGGGAGCAGACAGAGCCGAGGCTTGAGACCGAAG 2625  
 QY 861 GlyGlnGlnGlnSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro 880  
 DB 2626 GCGGAGGAGAGAGCTCAAGCGGCTCTTGGCCCTGTGTGACGCTCTCTGTTACACCTT 2685  
 QY 881 LeuThrAlaAlaGlnMetAlaProTrpMetLysArgLeuSerArgGlyGlnThrValGlu 900  
 DB 2686 CTGACCGCGCGGAGATGAGCCCTTACATGAACAGGCTTCCCGGGGCCAAGCGGTGAG 2745  
 QY 901 AspLeuLeuGlnValLeuSerAspIleAspGlnMetSerSerArgArgArgProGlnIleLeu 920  
 DB 2746 GATCTGCTGGAGGTTCTGTGATGACATGAGAGATGATCCCGGCGGAGCCAGATCTTG 2805  
 QY 921 SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGlnLysCysArgAsn 940  
 DB 2806 ACCTTCTTCTCCAGCACTGCTGAGCGGCTGATGAGTGGCGGAGAGAGTGTGCCGACAC 2865  
 QY 941 LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaAlaPhe 960  
 DB 2866 CTGCGCTTCAAGCTGCTGCCCTGCTGCTCATGACAGAACGCCAGACTTTCACCGCTTTC 2925  
 QY 961 LeuProThrPheMetTrpCysLeuGlnGlySerGlnAspPheGlnValValGlnThrAlaLeu 980  
 DB 2926 CTGCCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2985  
 QY 981 ArgAsnLeuProGlnIleTrpAlaLeuLeuCysGlnGlnHisAlaAlaValLeuLeuHisArg 1000  
 DB 2986 CGGAACCTGCTGAGTACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3045  
 QY 1001 AlaPheLeuValGlyMetTrpGlyGlnMetAspProSerAlaGlnIleSerGlnAlaLeu 1020  
 DB 3046 GCCTTCTGCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3105  
 QY 1021 ArgIleLeuHisMetGlnAlaValMet 1029  
 DB 3106 AGGATCTCTCATATGAGGCGGTGATG 3132

RESULT 6  
 US-10-032-990-21  
 Sequence 21, Application US/10032990  
 Publication No. US20030032060A1  
 GENERAL INFORMATION:  
 APPLICANT: Botstein, David  
 APPLICANT: Desnoyers, Luc  
 APPLICANT: Ferrara, Napoleone  
 APPLICANT: Fong, Sherman  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Pan, James  
 APPLICANT: Roy, Margaret Ann  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumes, Daniel  
 APPLICANT: Watanabe, Colin K.  
 APPLICANT: Wood, William I.





Db 1186 AGCAAGTCTCCAGGGCTGATCGAGTGAAGTCCCTCCAGCTGGAGGAGTGTGACT 1245  
 QY 401 AlapheSerSerAlaThrAlaAspAlaSerProPhaAlaCysLysProVal 420  
 Db 1246 GATTTCTTCTGACCACTGGAGTGTCTCTCCCGCTTTCAGCCGTGAAGCCCGTGTG 1305  
 QY 421 ValAlaSerSerLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 440  
 Db 1306 GGGGAGAGTCTCTGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1365  
 QY 441 AspGlyGlySerLeuGlnAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrp 460  
 Db 1366 GAGGGTGGAGAGCTGGAGAGCCCTGGCGGCTGGCGGCTCTAGAGGAGTGG 1425  
 QY 461 LeuGlnMetLeuAspProGlnValValSerSerCysProAspLeuGlnLeuArgLeu 480  
 Db 1426 CTGGAAATCTGGACCCAGAGTGTGCACAGCTGCCCCGAGCTGGAGTCAAGGCTGCTC 1485  
 QY 481 PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProTyrLeuLeu 500  
 Db 1486 TTCTCCGGAGAGAGGCAAGAGTCAAGGCTCCCTCGTCCGCTCCATACCTCTG 1545  
 QY 501 ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysTlaArgValLeuLeu 520  
 Db 1546 ACCCTTTCACCATCAGTCCAGCTGGCCACACTGACCAGTCCAGTCCGAGTCCGCTG 1605  
 QY 521 GlyLysSerArgGlnGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys 540  
 Db 1606 GCGAAGAGCCGGGAAACAGAGTTCACCCCTGCTCTGAGACTTCTCTGGGCTGCTG 1665  
 QY 541 LeuHisValProArgTleTrpGlnArgAspGlnArgTrpProGlnLysArgArgGln 560  
 Db 1666 ATCCATGTTCTCGCATCTGGAGAGGGGCGGAGACAGCAGCAGCCGCAAGAGCGGGAG 1725  
 QY 561 GlnLeuValLeuArgValGlnGlyProGlnLeuLleSerLeuValGlnLeuLleLeuAla 580  
 Db 1726 GAGCTGTGCTCTGGTCCAGGGCCCGGAGCTCATCAGCTGTGGAGTGTGATCCCTGGCC 1785  
 QY 581 GlnAlaGlnThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuLleGlnAlaArg 600  
 Db 1786 GAGGGGAGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1845  
 QY 601 LeuProLeuLeuLeuSerCysCysGlyAspAspGlySerValArgLysValThrGln 620  
 Db 1846 CTGGCCCTGCTGCTGAGTGTGCTGTGGGAGCAGTGAAGTGTGAGGAAGTGAAGGAG 1905  
 QY 621 HisLeuSerGlyCysLleGlnGlnTrpGlyAspSerValLeuGlnLysArgCysArgAsp 640  
 Db 1906 CACCTGTGAGGCTGCATCCAGAGTGGGAGACAGCTGTGGAGAGGGCTGCCAGAGAC 1965  
 QY 641 LeuLeuLeuGlnLeuTyrLeuGlnArgProGlnLeuArgValProValProGlnValLeu 660  
 Db 1966 CTTCCTCTCAGCTTACTTACAGCGGCGGAGCTCGGGTGTGCTGCTGAGTGTCTTA 2025  
 QY 661 LeuHisSerGlyGlyAlaAlaSerSerSerValCysLysLeuAspLysLeuLleHisArg 680  
 Db 2026 CTGCACAGGAGAGGGCTGCCAGACAGCAGCTGTCAAGTGTGAGAGGACTCATCCACCG 2085  
 QY 681 PheLleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnLysArgGlyAlaAsp 700  
 Db 2086 TTTCATCAGCTCTTGGCGAGACACAGCTCCCGGGCTGTGAAGAACGAGGGCGGAGT 2145  
 QY 701 AlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuLeuArgHisLeu 720  
 Db 2146 GCCAGCATGGCTGCGGAGAGCTGGGGTGGGCGACCCGCTGTGCTGTCAAGCACTG 2205  
 QY 721 PrometLleAlaAlaLeuLeuHisGlyArgTrpHisLeuAspPheGlnGlnPheArgGln 740  
 Db 2206 CCATGATGAGCGGCGCTCTGCAGGGCGGACACCACTCACTCCAGAGAGTTCGGGAG 2265  
 QY 741 GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuGlnProHis 760  
 Db 2266 CAGAACCACCTGAGCTTCTCTGACAGTGTGGGCTGTGAGAGCTGTGAGCGGCGAC 2325

QY 761 ValPheArgSerGlnHisGlnGlnValAlaLeuTrpAspCysLeuLeuSerPheLleArgLeu 780  
 Db 2326 GTTCTCCGAGGAGGAGACAGAGGGGCGCTGTGGAGCTGCTTCTGCTTCACTCCGCTG 2385  
 QY 781 LeuLeuAsnTyrArgLysSerSerArgHisLeuAlaAlaPheLleAsnLysPheValGln 800  
 Db 2386 CTGCTCAATTACAGGAAGTCTCCCGCATCTGCTGCTGCTTCAATCAAGAGTGTGAG 2445  
 QY 801 PheLleHisLysTyrLleThrTyrAsnAlaProAlaAlaLleSerPheLeuGlnLysHis 820  
 Db 2446 TTTCATCATAAGTACATTACTTCAATGAGCCCGAGCAGCATCTCTCTCTGCGAAGCGAC 2505  
 QY 821 AlaAspProLeuHisAspLeuSerPheAspAspSerAspLeuValMetLeuLysSerLeu 840  
 Db 2506 GCGAGCCGCTCCAGCACTGCTCTTGCAGACAGTGAAGCTGTGATGCTGAATCCCTC 2565  
 QY 841 LeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArgLysLeuAspGlnGln 860  
 Db 2566 CTTCGAGGGCTGAGCTGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2625  
 QY 861 GlyGlnGlnGlnSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro 880  
 Db 2626 GGGGAGAGAGAGAGCTCACCGGCTCTTGGCCCTGTGTCAGGCTCTCCCTGTACACCCCT 2685  
 QY 881 LeuThrAlaAlaGlnMetAlaProTyrMetLysArgLeuSerArgGlyGlnThrValGln 900  
 Db 2686 CTACAGCGCGCGAGATGGCCCTTACATGAACAGGCTTCCGGGGCCAAACGGTGGAG 2745  
 QY 901 AspLeuLeuGlnValLeuSerAspLleAspGlnMetSerArgArgArgProGlnLleLeu 920  
 Db 2746 GATCTCTGAGAGTGTGAGTGAATGACAGAGATGTCCCGGCGGAGACCCGAGATCTG 2805  
 QY 921 SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerLleGlnGlnCysArgAsp 940  
 Db 2806 AGCTTCTTCTCGACCACTGACGCGCGCTGATAGCTGGCCAGAGAGTGTCCCAAC 2865  
 QY 941 LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerLleAlaAlaPhe 960  
 Db 2866 CTGGCTTCAAGCTGTGGCCCTGCGCTCCATGCAACAGCCCAAGCTTTCAGCGCTTTC 2925  
 QY 961 LeuProThrPheMetTyrCysLeuGlySerGlnAspPheGlnValAlaGlnThrAlaLeu 980  
 Db 2926 CTGCCCCAGCTTCATGTACTGCTGCGGAGCAGCAGCATTTGAGTGTGTGCGAGCGCCCTC 2985  
 QY 981 ArgAspLeuProGlnTyrAlaLeuLeuCysGlnGlnHisAlaValLeuLeuHisArg 1000  
 Db 2986 CGGAACCTGCTGATGACGCTCTCTGTGCCAAGAGCAGCGGCTGTGCTGCCACCGG 3045  
 QY 1001 AlaPheLeuValGlyMetTyrGlyGlnMetAspProSerAlaGlnLleSerGlnAlaLeu 1020  
 Db 3046 GCTTCTGCTGTGGCATGTACGGCCAGATGAGACCCAGCGCGAGATCTCCGAGGCCCTG 3105  
 QY 1021 ArgLleLeuHisMetGlnAlaValMet 1029  
 Db 3106 AGGATCTGCAATGTGAGGCGCGTGAAGT 3132

RESULT 7  
 US-09-929-769-3  
 ; Sequence 3, Application US/09929769  
 ; Publication No. US20030055224A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gao, Wei-Oiang  
 ; APPLICANT: Polakis, Paul  
 ; APPLICANT: Shou, Jianyong  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Soriano, Robert  
 ; APPLICANT: Williams, P. Mickey  
 ; APPLICANT: Wu, Thomas D.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
 ; TREATMENT OF TUMOR  
 ; FILE REFERENCE: P5007R1-US

```

CURRENT APPLICATION NUMBER: US/09/929,769
CURRENT FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/089,653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/090,355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/104,257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/119,537
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/141,037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/162,506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/20111
PRIOR FILING DATE: 1999-09-01
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
PRIOR APPLICATION NUMBER: PCT/US00/00376
PRIOR FILING DATE: 2000-01-06
PRIOR APPLICATION NUMBER: PCT/US00/04342
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/13705
PRIOR FILING DATE: 2000-05-17
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/20118
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: 09/888,257
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 8
SEQ ID NO 3
LENGTH: 3437
TYPE: DNA
ORGANISM: Homo Sapien
US-09-929-769-3

Alignment Scores:
Pred. No.: 0 Length: 3437
Score: 5241.00 Matches: 1029
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-929-769-7 (1-1029) x US-09-929-769-3 (1-3437)
QY 1 MethisileuValIhIsAlaMetValIleuLeuThrlLeuGluProProAla 20
Db 46 ATGACATCTCTCGTGGTCCATGCTGATCTCTGACGCTGGGCGCCCTCGAGCC 105
QY 21 AspAspSerGluPheGlnAlaLeuAspIleTrpPheProGluGluProLeuPro 40
Db 106 GAGCAGCAGCAGTCCAGCGCTCTGACATCTGTTCCGAGGAGAGCCACTGCC 165
QY 41 ThAlaIheLeuValAspTrpSerGluGluAlaLeuLeuProAspTrpLeuLeu 60
Db 166 ACCGCTTCTCTGTGGCAGATCGAGAGCGCTCTCTGACTGCTGAAGCTG 225
QY 61 ArgMetIleArgSerGluValIleuArgLeuValAspAlaIleuGlnAspLeuGluPro 80
|||||
```

```

Db 226 CGCATGATCCGTTCTGAGTCTCCGCCCTGGTGAGCCGCCCTGCAGAGCTGAGCCG 285
QY 81 GlnGlnLeuLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerLyLeu 100
Db 286 CAGCAGCTGCTGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 345
QY 101 LeuGlnPheLeuAspGlnAlaValAlaIhIsAspProGlnTrhLeuGluGlnAsnIleMet 120
Db 346 CTCGAGTCTCTGAGCAGCAGGAGTGGCCAGACCCCGAGCTGAGCAGACATCATG 405
QY 121 AspLyAsnTrpMetAlaIhIsLeuValGlnIhIsGluuArgGlyAlaSerGlyLy 140
Db 406 GACAAGATTAATGAGCCCTGCTGAGTCCAGCATGAGCCGCGCTCCGAGGC 465
QY 141 GlnTrhPheIhIsSerLeuLeuThrlAserLeuProProArgAspSerThrluAla 160
Db 466 CAGACTTTCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
QY 161 ProLySProLySserSerProGluGlnProIleGlyGlnIArgValGlyThr 180
Db 526 CCAAGCAATTAATGAGCCCTGCTGAGTCCAGCATGAGCCGCGCTCCGAGGC 585
QY 181 GlnLeuArgValLeuGlyProGluAspAspLeuIleGlyMetPheLeuGlnIlePhePro 200
Db 586 CAGCTCCGAGTCTGAGCCCTGAGGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 645
QY 201 LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220
Db 646 CTCAGCCCGGAGCCCTCGGTGGCAGAGTCCAGTCCCGCCGCGGCTGCTGCTGCTG 705
QY 221 GlnAlaLeuGlyGlnGluLeuAlaArgValValGlnGlySerProGluValProGlyLy 240
Db 706 CAGCCCTGGGCGCAGAGCTGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765
QY 241 ThrValArgValLeuGlnAlaLeuAlaThrlLeuLeuSerSerProIhIsGlyValAlaLeu 260
Db 766 ACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
QY 261 ValMetSerMetIhIsArgSerIhIsPheLeuAlaCysProLeuLeuArgGlnLeuCysGln 280
Db 826 GTGATGTCATGACGACCGGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885
QY 281 TyrGlnArgCysValProGlnAspTrhGlyPheSerSerLeuPheLeuValLeuLeu 300
Db 886 TACAGCGCTGTGTGCCACAGACACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
QY 301 GlnMetLeuGlnTrpLeuAspSerProGlyValGlnGlyGlyProLeuArgAlaGlnLeu 320
Db 946 CAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005
QY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyLyLeu 340
Db 1006 AGGATGCTTGTGCCACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1065
QY 341 LeuArgLeuAlaGluAlaLeuAlaPheArgGlnAspLeuGluValIleSerSerThrlVal 360
Db 1066 CTGCGCTGCGCGAGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
QY 361 ArgAlaValIleAlaTrhLeuArgSerGlyGlnCysSerValGluProAspLeuIle 380
Db 1126 CGTCCGCTGATCGGACCTGAGTCTGGGAGAGCAGTGCAGCGGAGCCGACTGATC 1185
QY 381 SerLySValLeuGlnGlyLeuIleGluValArgSerProIhIsLeuGluGluLeuThr 400
Db 1186 AGCAAGTCTTCCAGGAGGCTGATGAGTGAGTCCGCCACCTGAGAGACTCTGACT 1245
QY 401 AlaPhePheSerAlaTrhAlaAspAlaIleSerProPheProAlaCysLySProValVal 420
Db 1246 GCATTCTTCTGTCCATGCGGAGTGTGCTCCCGCTTCCAGCTGTAAACCCGTTGTG 1305
QY 421 ValValSerSerLeuLeuGlnGluGlnIleGluProLeuAlaGlyLyLyProGlyVala 440
Db 1306 GTGCTGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1365
```



QY 441 AspGlySerLeuAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrp 460  
Db 1366 GAGGTGGACACCTGGAGCCGCTGGGGCCCTGCTGACGGCTCTCTACTGACAGC 1425  
QY 461 LeuGluMetLeuAspProGluValValSerSerCysProAspLeuGluLeuArgLeu 480  
Db 1426 CTGGAAATGGCTGGACCCCGAGGTGGTCAGACGCTGCCCGACCTCCAGCTCAGCTGCTC 1485  
QY 481 PheSerArgArgGlyGlyGlyGlnAlaGlnValProSerPheArgProTyrLeuLeu 500  
Db 1486 TTCCTCCGAGGAAGGGAAAGGTAGGCCAGGTGCCCTGGCTCCCTACTCCTCTG 1545  
QY 501 ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeuLeu 520  
Db 1546 ACCCTCTCAGACATCATCTCCAGCTGGGCCACACAGTCACAGATCGAGTCTGCTG 1605  
QY 521 GlyLysSerArgGluGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys 540  
Db 1606 GGCAGAGACCCGGGAACAGAGGTTCGACCCCTGCTGAGACTTCTCTGGGCGCTGC 1665  
QY 541 IleHisValProArgIleTrpGlnArgAspGlnArgTrpProGlnLysArgArgGlu 560  
Db 1666 ATCCATGTTCTCCGATCTGGCAGGGGGGAGCACAGCCACCCGAGAAAGCGCGGAG 1725  
QY 561 GluLeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGlnLeuLeuAla 580  
Db 1726 GAGCTGGGTGGGGCTCCAGGGCCCGGAGCTCATCAGCTGCTGAGCTGATCTGCGC 1785  
QY 581 GluAlaGluThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAlaArg 600  
Db 1786 GAGCGGAGAGCGCGAGGACAGAGCGGGACACAGCCGCTGACGCTCATCCAGGCGCG 1845  
QY 601 LeuProLeuLeuLeuSerCysCysCysGlyAspAspGluSerValArgLysValArgGlu 620  
Db 1846 CTGCCCCCTGCTCTAGCTGCTGCTGGGAGCATGAGTGAAGAGGTGACGGAG 1905  
QY 621 HisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArgAsp 640  
Db 1906 CACCTGTGAGGCTGCATCAGAGAGTGGGAGACAGCTGCTGGAAAGGCGTCCGAGAGC 1965  
QY 641 LeuLeuLeuGlnLeuTyrLeuGlnArgProGluLeuArgValProValProGluValLeu 660  
Db 1966 CTTCCTCTGAGCTTACTTACAGCGCGGAGCTGCGGGGTCCCGTGTGAGGCTCTTA 2025  
QY 661 LeuHisSerGluGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuLeuHisArg 680  
Db 2026 CTGCACAGCGAGGGGCTGCCAGCAGCAGCTCTCAAGCTGGAGGACTCATCCACCGC 2085  
QY 681 PheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAlaAsp 700  
Db 2086 TTCATCAGAGCTCTTGGCGAGACACAGCAGCTCCGGGCTTGGAAACCGAGGGCGGAT 2145  
QY 701 AlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuArgHisLeu 720  
Db 2146 GCCACATAGGCTTGGCGAAGCTGGCGGTGGCGCACCCGCTCTGCTCAGGACCTG 2205  
QY 721 PrometIleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPheGlnGluPheArgGln 740  
Db 2206 CCCATGATCGGGCTCTCTGACAGCGCGCACCTCACTTCCAGAGATTCCGGCGAG 2265  
QY 741 GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuLeuProHis 760  
Db 2266 CAGAACACACCTAGCTGCTCTCTGACAGCTGGGGCTGCTGAGACTGCTGACGCGGAC 2325  
QY 761 ValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgLeu 780  
Db 2326 GTGTTCCCGACGAGCAGACAGGGGGCTGGAGACTGCTCTCTCATCCGCGCTG 2385  
QY 781 LeuLeuAsnTyrArgLysSerSerArgHisLeuAlaAlaPheIleAsnLysPheValGln 800  
Db 2386 CTGCTGAATTACAGGAAGTCTCCCGCATCTGGCTGCTTCATCAACAAAGTTGTGAGC 2445

QY 801 PheIleHisLysTyrIleThrTyrAsnAlaProAlaAlaIleSerPheLeuGlnLysHis 820  
Db 2446 TTCATCCATTAAGTACATTACCTACCAATGCCCCAGCAGCATCTCTCTCTGAGAAAGCAGC 2505  
QY 821 AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu 840  
Db 2506 GCCGACCCGCTCCACAGACTCTCTCTTCGACAAACACTGACCTGTGATCTGAAATCCCTC 2565  
QY 841 LeuAlaGlyLeuSerLeuProSerArgAspAspArgTrpAspArgGlyLeuAspGlyGlu 860  
Db 2566 CTTCGAGGGCTCAGCTCTGCCACGACGAGGACGACAGGACCGACCGAGGCTTGACCAAG 2625  
QY 861 GlyGluGluGluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro 880  
Db 2626 GCGGAGAGAGAGAGCTCAGCGGCTCTGCCCCCTGGTCAGAGTCTCTCTGTTACCCCT 2685  
QY 881 LeuThrAlaAlaGluMetAlaProTyrMetLysArgLeuSerArgGlyGlnThrValGlu 900  
Db 2686 CTGACCGGGCGGAGATGAGCCCTTCATGAAACGGCTTCCCGGGGCCAAACGGGTGAG 2745  
QY 901 AspleuLeuGluValLeuSerAspIleAspGluMetSerArgArgArgProGluLeuLeu 920  
Db 2746 GATCTGCTGAGAGTCTGATGATACATACAGAGATGTCCTCCGGGAGAGACCGAGTCTG 2805  
QY 921 SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGluGluCysArgAsn 940  
Db 2806 AGCTTCTCTCTGACCACTGACGCGGCTGATGAGACTCGGGCGAGAGTGGCCGCAAC 2865  
QY 941 LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaPhe 960  
Db 2866 CTGCGCTTCAGCTGCGCTGCGCTTCATGAGAACAGCGCCAGCATTTGACACCCCTTTC 2925  
QY 961 LeuProThrPheMetTyrCysLeuGlySerGlnAspPheGluValGlnThrAlaLeu 980  
Db 2926 CTGCCACGTTATATGATCTGCTGGGACAGCCAGGACTTTGAGAGTGTGACAGCGCCCTC 2985  
QY 981 ArgAsnLeuProGluTyrAlaLeuLeuCysGlnGluHisAlaAlaValLeuLeuHisArg 1000  
Db 2986 CGGAACCTGCGTAGTACGCTCTCTGTCGCAAGACAGCGGCTGTGCTCCACCGG 3045  
QY 1001 AlaPheLeuValGlyMetTyrGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeu 1020  
Db 3046 GCCTTCTGCTGGGATGATGACGCGCAGATGAGACCCAGCGCAGATCTCCGAGGCGCTG 3105  
QY 1021 ArgIleLeuHisMetGlnAlaValMet 1029  
Db 3106 AGGATCTGCAATGAGAGCGGTGATG 3132

RESULT 8  
US-10-032-996-21  
Sequence 21, Application US/10032996  
Publication No. US20030054447A1  
GENERAL INFORMATION:  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Geo, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2930R1C3  
CURRENT FILING DATE: US/10/032,996  
CURRENT APPLICATION NUMBER: 60/095,325  
PRIOR APPLICATION NUMBER: 1998-08-04

```

? PRIOR APPLICATION NUMBER: 60/112,851
? PRIOR FILING DATE: 1998-12-16
? PRIOR APPLICATION NUMBER: 60/113,145
? PRIOR FILING DATE: 1998-12-16
? PRIOR APPLICATION NUMBER: 60/113,511
? PRIOR FILING DATE: 1998-12-22
? PRIOR APPLICATION NUMBER: 60/115,558
? PRIOR FILING DATE: 1999-01-12
? PRIOR APPLICATION NUMBER: 60/115,565
? PRIOR FILING DATE: 1999-01-12
? PRIOR APPLICATION NUMBER: 60/115,733
? PRIOR FILING DATE: 1999-01-12
? PRIOR APPLICATION NUMBER: 60/119,341
? PRIOR FILING DATE: 1999-02-09
? PRIOR APPLICATION NUMBER: 60/119,537
? PRIOR FILING DATE: 1999-02-10
? PRIOR APPLICATION NUMBER: 60/119,965
? PRIOR FILING DATE: 1999-02-12
? PRIOR APPLICATION NUMBER: 60/162,506
? PRIOR FILING DATE: 1999-10-29
? PRIOR APPLICATION NUMBER: 60/170,262
? PRIOR FILING DATE: 1999-12-09
? PRIOR APPLICATION NUMBER: 60/187,202
? PRIOR FILING DATE: 2000-03-03
? PRIOR APPLICATION NUMBER: PCT/US99/12252
? PRIOR FILING DATE: 1999-06-02
? PRIOR APPLICATION NUMBER: PCT/US99/28634
? PRIOR FILING DATE: 1999-12-01
? PRIOR APPLICATION NUMBER: PCT/US99/28551
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US00/03565
? PRIOR FILING DATE: 2000-02-11
? PRIOR APPLICATION NUMBER: PCT/US00/04414
? PRIOR FILING DATE: 2000-02-22
? PRIOR APPLICATION NUMBER: PCT/US00/05841
? PRIOR FILING DATE: 2000-03-02
? PRIOR APPLICATION NUMBER: PCT/US00/08439
? PRIOR FILING DATE: 2000-03-30
? PRIOR APPLICATION NUMBER: PCT/US00/14941
? PRIOR FILING DATE: 2000-05-30
? PRIOR APPLICATION NUMBER: PCT/US00/15264
? PRIOR FILING DATE: 2000-06-02
? PRIOR APPLICATION NUMBER: PCT/US00/32678
? PRIOR FILING DATE: 2000-12-01
? NUMBER OF SEQ ID NOS: 38
? SEQ ID NO 21
? LENGTH: 3437
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-10-032-996-21

```

```

Alignment Scores:
Prid. No.: 0 Length: 3437
Score: 5241.00 Matches: 1029
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Gaps: 0
DB: 9

```

US-09-929-769-7 (1-1029) x US-10-032-996-21 (1-3437)

```

QY 1 MeHsIleuValValHsAlaMeValIleuLeuHrLeuGluYrProRoArgAla 20
Db 46 ATCCACATCTCTGTGGCCATGTCATCTCTGTGGCCGCTCGAGCC 105
QY 21 AspAspSerGluPheGlnAlaLeuAspIleTrpPheProGluGluYrProLeuPro 40
Db 106 GAGACAGCAGGATTCAGGGGCTGCTGACATCTGTTCGAGAGAGAGCCACTGCCC 165
QY 41 ThAlaPheLeuValAspThrSerGluGluAlaLeuLeuProAspTrpLeuYsLeu 60
Db 166 ACCGCTTCTCTGTGGACATCGAGAGAGGCGCTGCTCTCTGACTGAGAGCTG 225

```

```

QY 61 ArgMetIleArgSerGluValIleuArgLeuValAspAlaAlaLeuGlnAspLeuGluPro 80
Db 226 CGCATGATCCGCTTCTGAGGTCGCTCCGCTGTCGAGCCGCCCTGAGAGCCGTCGAGACC 285
QY 81 GlnGlnLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerYsLeu 100
Db 286 CAGAGGCTGCTGTGTGTCGAGTCGCTGTTCGATCCCTGTCACATGAGCAAACTC 345
QY 101 LeuGlnPheLeuAspGlnAlaValAlaHsAspProGlnHrLeuGluGlnAsnIleMet 120
Db 346 CTCAGTTCCTGAGCCAGGAGGAGTGGCCAGACCCCGAGACTCGGAGAGAACATCATG 405
QY 121 AspYsAsnTrpMetAlaHsLeuValGluValGlnHsGluArgGlyAlaSerGly 140
Db 406 GACAGAATTCATATGCGCCCTGTCGAGTCCAGATAGAGCGCCGCTCCGAGGCG 465
QY 141 GlnThrPheHsSerLeuLeuThrAlaSerLeuProRoArgTrpAspSerThrGluAla 160
Db 466 CAGACTTCCACTCTGTGTCACAGCCTCCCTGCGCCGCGGAGACACACAGAGGCA 525
QY 161 ProYsProYsSerSerProGluGlnProIleGlyGlnArgIleArgValGlyThr 180
Db 526 CCCAAACCAAGAGACAGCCAGAGACCCATAGAGCCAGGCGGAGTTGGGGAGACC 585
QY 181 GlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200
Db 586 CAGCTCCGGGTGCTGGGCTCGTAGAGACCTGCTGCTGCTCCATGTTCTCCAGATTTCGCC 645
QY 201 LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220
Db 646 CTCAGCCCGGAGCCCTGCTGTCAGAGCTCCAGTCCCGCCGCTGCTGCTGCTGCTGCTG 705
QY 221 GlnAlaLeuGlyGlnLeuLeuAlaArgValValGlnYsSerProGluValProGlyIle 240
Db 706 CAGGCTCTGGGAGAGAGTGGCCGCTGTCAGAGGAGCCGCGAGGAGTGGCGGCGATC 765
QY 241 ThrValArgValLeuGlnAlaAlaThrLeuLeuSerSerProHsGlyValAlaLeu 260
Db 766 ACGGTGCTGCTCTGAGGCTCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
QY 261 ValMetSerMetHsArgSerHsPheLeuAlaCysProLeuLeuArgGlnLeuCysGln 280
Db 826 GTGATGTCATGACCGTACCGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885
QY 281 TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuYsValLeuLeu 300
Db 886 TACCAGGCTGTGTGTCACAGACACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
QY 301 GlnMetLeuGlnHrPheLeuAspSerProGlyValGlnGlyProLeuArgAlaGlnLeu 320
Db 946 CAGATGCTGAGTGGCTGTCAGAGCCCTGCTGAGAGGCGGCGCCCTGCGGACACTC 1005
QY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeu 340
Db 1006 AGGATGCTTCCAGCCAGGCTGAGGCTGAGGAGACAGGCTGAGGAGGAGGAGGCTG 1065
QY 341 LeuArgLeuAlaGluAlaLeuAlaPheArgGlnAspLeuGluValValSerSerThrVal 360
Db 1066 CTGCGCTGCGCAGAGGCTGAGGCTTCTGTCAGAGACTGAGAGTGGTCACTCCACGCTC 1125
QY 361 ArgAlaValIleAlaThrLeuArgSerGlyGluGlnCysSerValGluProAspLeuIle 380
Db 1126 CGTGGCTGATCGCACCCCTGAGGCTGAGGAGACAGGCTGAGGAGGAGGAGGAGGCTG 1185
QY 381 SerYsValLeuGlnGlyLeuIleGluValArgSerProHsSerLeuGlnLeuLeuThr 400
Db 1186 AGCAAACTCTCCAGGAGGCTGATCGAGTGAAGTCCCGCCAGCTGAGAGAGTGGTACT 1245
QY 401 AlaPhePheSerAlaThrAlaAspAlaAlaSerProPheProAlaCysYsProValVal 420
Db 1246 GCATTTCTTCTGCTCAGCTGCGAGTGGCTCCCGCTTCTCAGACCTGTAGAGCCGTTGTC 1305
QY 421 ValValSerSerLeuLeuGlnGluGluProLeuAlaGlyYsProGlyAla 440

```

```

|||||
Db 1306 GTGGAGACCTCCCTGCTCTGAGAGAGAGAGCCCTGCTGGGGAGAGCCGGCTCG 1365
Qy 441 ASpglySerLeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrp 460
Db 1366 GACGGTGCACCTGGAGAGCCGGTGGGCTGCTGACAGGCTCTCTAGTGGACGTGG 1425
Qy 461 LeuGluMetLeuAspProGluValValSerSerCysProAspLeuGluLeuArgLeuLeu 480
Db 1426 CTGGAATAGCTGAGACCCGGAGGTGTACAGAGCTCCCGGACCTCAGCTCAGGCTGCTC 1485
Qy 481 PheSerArgArgGlyGlyGlyGlyGlnAlaGlnValProSerPheArgProTyrLeuLeu 500
Db 1486 TTCTCCCGAGAGAGGGGCAAGAGTCAGGCCAGGTGCCCTGCTCCCTACCTCTCTG 1545
Qy 501 ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeuLeu 520
Db 1546 ACCCTCTTCACGACATCTCCAGCTGGGCCACACTGCACAGTGCATCCGAGTCTCTGCTG 1605
Qy 521 GlySerSerArgGluGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys 540
Db 1606 GGCAAGACCCGGGAACAGAGGTGCACCCCTGCTCTGTGACTTCTCTGGGCTGCTG 1665
Qy 541 IleHisValProArgIleTrpGlnGlyArgAspGlnArgTrpProGlnLysArgArgGlu 560
Db 1666 ATCCATGTCCTCGCATCTGGCAGGGCGGAGCCAGCCAGCCCGCAGAAAGCGCGGAG 1725
Qy 561 GluLeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGluLeuIleLeuAla 580
Db 1726 GAGCTGGTGTGGGGGTCCAGGGCCGAGCTCATCAGCTGGGTGGAGTCTGCTGGCC 1785
Qy 581 GluAlaGluThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAlaArg 600
Db 1786 GAGGGGAGAGAGCGGAGAGCCAGAGCCGAGACAGCCGCTCAGCTCATCAGGCCCGG 1845
Qy 601 LeuProLeuLeuLeuSerCysCysGlyAspAspGluSerValArgLysValThrGlu 620
Db 1846 CTGCCCCCTGCTCAGCTGTGCTGTGGGACGATGAGAGTGTGAGAGGTGACGAG 1905
Qy 621 HisLeuSerGlyCysIleGlnGlnTrpGlnAspSerValIleuGlyArgArgCysArgAsp 640
Db 1906 CACCTGTAGGCTGCATTCAGAGAGTGGGAGACAGCTGTGGGAGAGCGCTGCGGAGAC 1965
Qy 641 LeuLeuLeuGlnLeuTyrLeuGlnArgProGluLeuArgValProValProGluValLeu 660
Db 1966 CTTCCTCTGACGCTCTACTACAGAGCGGAGCGGAGCGGAGCGCTGCTGAGGTCTCA 2025
Qy 661 LeuHisSerGluGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuIleHisArg 680
Db 2026 CTGCACACCGAAGGGGCTGCCAGCAGCAGCGCTGCCAAGCTGGAGGAGCTCATCCACCGC 2085
Qy 681 PheIleThrLeuLeuAlaAspThrSerAspSerArgAlaIleuGluAsnArgGlyAlaAsp 700
Db 2086 TTTCATCAGCTCTTCTGGGAGACACAGCAGCTCCCGGGCTTGGAGAACCGAGGGCGGAT 2145
Qy 701 AlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuArgHisLeu 720
Db 2146 GCCAGACATGGCTGGCGGAGGTGGCGGTGGGAGCCCGCTGCTGCTGACGACCTG 2205
Qy 721 PrometIleAlaIleAlaLeuLeuHisGlyArgThrHisLeuAsnPheGlnGluPheArgGln 740
Db 2206 CCCATGATCGGGGCTCTCTCAGCGCGCAGCCACCTCAACTTCCAGAGAGTTCGGGAG 2265
Qy 741 GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuGlnProHis 760
Db 2266 CAGAAACCACTGAGCTGCTTCTGACAGCTGGGGCTGCTGAGAGCTGCTGACGACCGGAC 2325
Qy 761 ValPheArgSerGluHisGlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgLeu 780
Db 2326 GTGTTCACGACGAGCAGCAGGAGGGCGCTGTGGAGACCTCTTCTGCTTATCCGGCTG 2385
Qy 781 LeuLeuAsnTyrArgLysSerSerArgHisLeuAlaIlePheIleAsnLysPheValGln 800
|||||

```

```

Db 2386 CTGCTGAATTACAGGAAGTCTCCCGCATCTGCTGCTTCATCAACAAGTTTGTGCAG 2445
Qy 801 PheIleHisLysTyrIleThrTyrAsnAlaProAlaIleIleSerPheLeuGlnLysHis 820
Db 2446 TTTCATCTCAATAGTACATTACCTACAAATGCCAGAGACCATCTCTTCTGACAGAGAC 2505
Qy 821 AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu 840
Db 2506 GCCGACCCGCTCCAGACCTGTCTTCGACAAACATGACCTGATGATCTGTAATCCCTC 2565
Qy 841 LeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArgGlyLeuAspGlu 860
Db 2566 CTTCGAGGGCTCAGCTCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2625
Qy 861 GlyGluGluGluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro 880
Db 2626 GCGGAGAGAGAGCTCAGCCGGCTCTTGGCTGTGAGGTGTCTCTGCTGCTGCTGCT 2685
Qy 881 LeuThrAlaAlaGluMetAlaProTyrMetLysArgLeuSerArgGlyIleThrValGlu 900
Db 2686 CTGACCGCGGCGGAGATGGCCCTACATGAAGCGCTTCCCGGGGCAACGGGTGGAG 2745
Qy 901 AspLeuLeuGluValLeuSerAspIleAspGluMetSerArgArgArgProGluIleLeu 920
Db 2746 GATCTGTGTGAGGTCTCTAGTGCATAGACGAGATGTCCCGCGGAGACCGCAGATCTCTG 2805
Qy 921 SerPhePheSerThrAsnLeuGlnArgLeuLeuMetSerSerAlaGluLysCysArgAsn 940
Db 2806 AGCTTCTTCTGCACCACTGACGCGCTGAGAGCTGAGAGCTGGCGGAGAGGTGGCGCAAC 2865
Qy 941 LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaIlePhe 960
Db 2866 CTGCGCTTCAGCTGTGGCTGTGCTGCATGACGAGAACAGCCGAGATTGACGCCCTTTC 2925
Qy 961 LeuProThrPheMetTyrCysLeuGlySerGlnAspPheGluValGlnIleThrAlaLeu 980
Db 2926 CTGCCCCAGCTTCATATGCTGCTGGCGCAGCCAGGACTTGTGAGAGTGTGTGACAGCGCCCTC 2985
Qy 981 ArgAsnLeuProGluTyrAlaLeuLeuLeuCysGlnGlnIleAlaAlaValLeuLeuHisArg 1000
Db 2986 CGGAACCTGCTGAGTACGCTTCTGTGCCAAGAGCAGCGGCTGTGCTGTCCACCGG 3045
Qy 1001 AlaPheLeuValGlyMetTyrGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeu 1020
Db 3046 GCCTTCTGTGTGGGATGATGAGGCGCAGATGAGACCCAGGCGCAGATCTCGAGGCCCTG 3105
Qy 1021 ArgIleLeuHisMetGluAlaValMet 1029
Db 3106 AGGATCTTCATATGTGAGGCGGTGATG 3132

RESULT 9
US-10-033-396-21
; Sequence 21, Application US/10033396
; Publication No. US20030077657A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2930R1C4
; CURRENT FILING DATE: 2001-12-27

```

```

? PRIOR APPLICATION NUMBER: 60/095,325
? PRIOR FILING DATE: 1998-08-04
? PRIOR APPLICATION NUMBER: 60/112,851
? PRIOR FILING DATE: 1998-12-16
? PRIOR APPLICATION NUMBER: 60/113,145
? PRIOR FILING DATE: 1998-12-16
? PRIOR APPLICATION NUMBER: 60/113,511
? PRIOR FILING DATE: 1998-12-22
? PRIOR APPLICATION NUMBER: 60/115,558
? PRIOR FILING DATE: 1999-01-12
? PRIOR APPLICATION NUMBER: 60/115,565
? PRIOR FILING DATE: 1998-01-12
? PRIOR APPLICATION NUMBER: 60/115,733
? PRIOR FILING DATE: 1999-01-12
? PRIOR APPLICATION NUMBER: 60/119,341
? PRIOR FILING DATE: 1999-02-09
? PRIOR APPLICATION NUMBER: 60/119,537
? PRIOR FILING DATE: 1999-02-10
? PRIOR APPLICATION NUMBER: 60/119,965
? PRIOR FILING DATE: 1999-02-12
? PRIOR APPLICATION NUMBER: 60/162,506
? PRIOR FILING DATE: 1999-10-29
? PRIOR APPLICATION NUMBER: 60/170,262
? PRIOR FILING DATE: 1999-12-09
? PRIOR APPLICATION NUMBER: 60/187,202
? PRIOR FILING DATE: 2000-03-03
? PRIOR APPLICATION NUMBER: PCT/US99/12252
? PRIOR FILING DATE: 1999-06-02
? PRIOR APPLICATION NUMBER: PCT/US99/28634
? PRIOR FILING DATE: 1999-12-01
? PRIOR APPLICATION NUMBER: PCT/US99/28551
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US00/03565
? PRIOR FILING DATE: 2000-02-11
? PRIOR APPLICATION NUMBER: PCT/US00/04414
? PRIOR FILING DATE: 2000-02-22
? PRIOR APPLICATION NUMBER: PCT/US00/05841
? PRIOR FILING DATE: 2000-03-02
? PRIOR APPLICATION NUMBER: PCT/US00/08439
? PRIOR FILING DATE: 2000-03-30
? PRIOR APPLICATION NUMBER: PCT/US00/14941
? PRIOR FILING DATE: 2000-05-30
? PRIOR APPLICATION NUMBER: PCT/US00/15264
? PRIOR FILING DATE: 2000-06-02
? PRIOR APPLICATION NUMBER: PCT/US00/32678
? PRIOR FILING DATE: 2000-12-01
? NUMBER OF SEQ ID NOS: 38
? SEQ ID NO 21
? LENGTH: 3437
? TYPE: DNA
? ORGANISM: Homo sapiens
US-10-033-396-21

Alignment Scores:
Pred. No.: 0 Length: 3437
Score: 5241.00 Matches: 1029
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-929-769-7 (1-1029) x US-10-033-396-21 (1-3437)
OY 1 Methis11leuValValHisAlaMetVal11leuLeuThrLeuGlyProProArgAla 20
Db 46 ATGCACATCTCTGCTGCATGCATGCATGATCTCTGTCGCGCCGCGCTCCAGCC 105
OY 21 ASPASerGluPheGlnAlaLeuLeuAsp11eTrrPheProGluGluysProLeuPro 40
Db 106 GACGACAGGAGTTCACAGGCGCTGTCGACATCTGCTTCGCGAGAGAACGACATGCC 165
OY 41 AlaPheLeuValAspTrsSerGluGluAlaLeuLeuProAspTrsPheLeuysLeu 60

```

```

Db 166 ACCGCTCTCTGCTGCACATCGAGAGGCGCTCTCTCTGACTGCGTGAAGCTG 225
OY 61 ArgMet11earSerGluValLeuArgValAspAlaAlaLeuGlnAspLeuGluPro 80
Db 226 CGCATATCTCGTTCTGAGGTGCTCCGCTGCTGAGCGCGCTGAGGACCTCGAGCGCG 285
OY 81 GlnGlnLeuLeuPheValGlnSerPheGly11eProValSerSerMetSerLysLeu 100
Db 286 CAGCAGCTGCTGCTGCTGCTGAGTGGATCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 345
OY 101 LeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGlnAsn11eMet 120
Db 346 CTCAGATCTCTGAGCAGGAGGAGTGGCCAGACCCGCTGCTGCTGCTGCTGCTGCTGCT 405
OY 121 AspLysAsnTyrMetAlaHisLeuValGluValGlnHisGluArgGly11aSerGlyGly 140
Db 406 GACAGAGATTACATGGCCACCGCTGAGAGTTCACAGATGAGGCGCGCTCCGAGGCG 465
OY 141 GlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgArgAspSerThrGluAla 160
Db 466 CAGACTTCTCCACTCTTGTCTCACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
OY 161 ProLysProLysSerSerProGluGlnProLysGlnGlyGlnGlyArg11eArgValGlyThr 180
Db 526 CCCAAGCCAAAGACAGCCAGCCAGACCCATAGCCAGGCGCGGATTCGGGTGGGAGCC 585
OY 181 GlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGln11ePhePro 200
Db 586 CAGCTCCGGGTGCTGAGGCGCTGAGGAGCAGCTGCTGAGGATGCTTCCAGATTTCGCCG 645
OY 201 LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220
Db 646 CTCAGCGCCGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705
OY 221 GlnAlaLeuGlnGlnLeuAlaArgValAlaGlnGlySerProGluValProGly11e 240
Db 706 CAGGCGCTGAGCCAGGAGAGCTGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765
OY 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGly11aLeu 260
Db 766 ACGGTCTGCTCTGCTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
OY 261 ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuGlnGln 280
Db 826 GTGATGTTCATGACCGTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 885
OY 281 TyrGlnArgCysValProGlnAspTrnGlyPheSerSerLeuPheLeuysValLeuLeu 300
Db 886 TACACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 945
OY 301 GlnMetLeuGlnTrpLeuAspSerProGlyValGlnGlyGlyProLeuArgAlaGlnLeu 320
Db 946 CAGATCTGAGTGGCTGAGGCTGAGCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005
OY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeu 340
Db 1006 AGGATGCTTGGCCAGGCGCTGAGCGGCGGAGGCTGAGTATGTGCGAGGGGGGCTC 1065
OY 341 LeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValAlaSerSerThrVal 360
Db 1066 CTGCGCTGAGCCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
OY 361 ArgAlaVal11leuAlaThrLeuArgSerGlyGlnGlnCysSerValGlnProAspLeu11e 380
Db 1126 CGTGGCGGTATGCCACCGCTGAGGCTGAGGAGACACTGACGAGTGGAGCGGACCTGAGTC 1185
OY 381 SerLysValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 400
Db 1186 AGCAAAAGTCTCCAGGCGGTGATGAGGTGAGGTCCCGCCAGCTGAGAGAGCTGCTGACT 1245
OY 401 AlaPhePheSerL1aThrAlaAspAla11aSerProPheProAlaCysLysProVal11a 420
Db 1246 GCATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305

```



[illegible]

Qy	41	ThAlaPheLeuValAspThrSerGluGluAlaLeuLeuLeuProAspTrpLeuVal	60
Db	166	ACC GCCCTTCGGTGGACACATCGAGAGAGCCGTGCTTCTTCACTAGCTGAAGCTG	225
Qy	61	ArgMetIleArgSerGluValLeuArgLeuValAspAlaAlaLeuGlnAspLeuGluPro	80
Db	226	CGCATGATCCGCTTCTGAGAGTCTCCGCGCTGGTGGACCCGCTCTGCAGACCTCGAGCCG	285
Qy	81	GlnGlnLeuLeuLeuPheValGlnSerPheGlyTlPheProAspSerMetSerIleu	100
Db	286	CAGACGCTGGCTGGTTCGTGCACTCTTTGGCATCCCGCTGCACATGACCAACTC	345
Qy	101	LeuGlnPheLeuAspGlnAlaValAlaIleAspProGlnThrLeuGluGlnAsnIleMet	120
Db	346	CTCCAGTTCCTGGACACAGGACGTGGCCACAGACCCCGACACTGTGGAGCAACATCATG	405
Qy	121	AspLysAsnTrpMetAlaIleHisLeuValGluValGlnHisGluArgIleAspSerGlyIle	140
Db	406	GACAAAGATTACATGGGCCACACTGGTGTGAGTCCAGCATAGCGCGCGCTCCGGAGGC	465
Qy	141	GlnTrpPheHisSerLeuLeuThrAlaSerLeuProProArgTrpAspSerThrGluAla	160
Db	466	CAGACTTTCACCTCTTGCTACAGCTCTCCCTGCGCCCGCGGACACACAGAGCA	525
Qy	161	ProLysProLysSerSerProGluGlnProIleGlyGlnGlyArgIleArgIleGlyThr	180
Db	526	CCCAACCCAAAGACACAGCCCGACACACCCATAGCCAGCGAGCGCGATCCGGTGGGAGCC	585
Qy	181	GlnLeuArgValIleuGlyProGlnAspAspLeuAlaGlyMetPheLeuGlnIlePhePro	200
Db	586	CAGCTCGGGGTGGGGCCCTGAGAGACACTGGCTGGCATGTCTCCAGATTTTCCGG	645
Qy	201	LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln	220
Db	646	CTCAGCCCGGACCTCGGTGGCAAGCTCCAGTCCCGCCCGGTGGGCCCTCGCCCTCGAG	705
Qy	221	GlnAlaLeuGlyGlnGluLeuAlaArgValValGlnGlySerProGluValProGlyIle	240
Db	706	CAGGCCCTGGGCCAGAGAGCTGGCCGGCTGCTCCAGGGCAGCCCGAGGGTGGCCATC	765
Qy	241	ThrValArgValIleuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyAlaLeu	260
Db	766	ACGGGCGGTGTCGAGAGGCCCTGCCACCTGTGACGTCCCAACAGCGGGTGGCTCG	825
Qy	261	ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgIleuLeuGlyGln	280
Db	826	GTGATGTCACATGACCCGTAGCACCTTCTGGCCCTGCCGTGTGGCCAGCTCTCCAG	885
Qy	281	TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValIleuLeu	300
Db	886	TACCAAGCGCTGTGGCCACAGACACCGGCTTCTCTCGCTTCTCCGAAGGTCTCTG	945
Qy	301	GlnMetLeuGlnTrpLeuAspSerProGlyValGluGlyIleProLeuArgAlaGlnLeu	320
Db	946	CAGATGCTGAGTGGCTGGAGACGCCGTGGGTGAGAGGGCGGCCCTCTGCGGCACAGCTC	1000
Qy	321	ArgMetLeuAlaSerGlnAlaSerAlaGlyValArgArgLeuSerAspValArgGlyGlyLeu	340
Db	1006	AGGATGCTGGCACCGACTCAGCCGGCGGCGACGCTCATGATGTCCAGAGGGGGCTC	1060
Qy	341	LeuArgLeuAlaGluAlaLeuAlaPheArgGlnAspLeuGluValValSerSerThrVal	360
Db	1066	CTGGCGCTGGCCGAGGGCCGTGGCTTCGTCAGGACTGTGAGGNGNCACCTCCACCGTC	1120
Qy	361	ArgAlaValIleAlaThrLeuArgSerGlyGluGlnCysSerValGluProAspLeuIle	380
Db	1126	CGTGCCGTCAATCGCCACCTGAGTCTGGGGAGCAGGTGAGCGGTGAGCGGACCTGATC	1180
Qy	381	SerIysValIleuGlnGlyLeuIleGluValArgSerProHisLeuGluGlnLeuLeuThr	400
Db	1186	AGCAAAAGCTCTCCAGGGGGTATGTAGAGTGAAGTCCCCACCTGTGAGAGACGTGCTACT	1240
Qy	401	AlaPhePheSerAlaThrAlaAspAlaAlaSerProPheProAlaCysLysProValVal	420

```

|||||
Db 1246 GCATTCCTCTGTCACAGGATGCTGCTCCCGTTCCAGCTGTAAAGCCGCTGG 1305
|||
Qy 421 ValValSerSerLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 440
Db 1306 GTGGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1365
Qy 441 AspGlySerLeuGlnAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrp 460
Db 1366 GACGGTGGCAGCTGGAGCCGTGGGCGCTGGGCGCTGTCAAGCCCTCTAGTGAAGTGG 1425
Qy 461 LeuGlnMetLeuAspProGlnValValSerSerCysProAspLeuGlnLeuArgLeu 480
Db 1426 CTGGAAAGCTGGACCCCGAGGTGGTCAAGCTGCCCGACCTCACTCAGGCTGCTC 1485
Qy 481 PheSerArgArgGlyGlyGlyGlnAlaGlnValProSerPheArgProTrpLeuLeu 500
Db 1486 TTCTCCCGAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1545
Qy 501 ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeu 520
Db 1546 ACCCTCTCAGCAGCATCATCAGCTGGGCCACACTGCACAGTGCATCCGAGTCTGCTG 1605
Qy 521 GlySerSerArgGlnGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys 540
Db 1606 GGCAGAGACCCGGGAACAGAGGTTGACCCCTCTGACTTGTGACTCTCTGAGGCTGAC 1665
Qy 541 IleHisValProArgIleTrpGlnGlyArgGlnArgGlnArgGlnArgGlnArgGln 560
Db 1666 ATCCAGTCTCTCGCATGTGGAGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1725
Qy 561 GlnLeuValLeuArgValGlnGlyProGlnLeuLeuSerLeuValGlnLeuLeuAla 580
Db 1726 GAGCTGGTGTGGGGGCTCAGAGGCCGGAGCTCATCAGCTGGTGGAGTGAATCCTGGCC 1785
Qy 581 GlnAlaGlnTrpArgSerGlnAspGlyAspPheAlaAlaCysSerLeuIleGlnAlaArg 600
Db 1786 GAGGGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1845
Qy 601 LeuProLeuLeuLeuSerCysCysGlyAspAspGlnSerValArgLeuValThrGln 620
Db 1846 CTGCCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1905
Qy 621 HisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArgAsp 640
Db 1906 CACCTGTGAGCTGCATCAGAGAGTGGGAGAGACAGCTGCTGGAGGGGCTGCCGAGAGC 1965
Qy 641 LeuLeuLeuGlnLeuTrpLeuGlnArgProGlnLeuArgValProValProGlnValLeu 660
Db 1966 CTTCCTCTGAGCTCTACCTACAGGGGCGGAGCTGCGGGTCCCTGCTCTGAGGCTCTA 2025
Qy 661 LeuHisSerGlnGlyAlaAlaSerSerSerValCysIleValAspGlyLeuLeuHisArg 680
Db 2026 CTGCACAGCAGAGGGGCTGCCAGCAGACGCTCTCAAGCTGGAGGAGCTCATCCACGCC 2085
Qy 681 PheIleThrLeuLeuAlaAspTrpHisSerSerArgAlaLeuGlnAsnArgGlyAlaAsp 700
Db 2086 TTTCATCAGCTCTCTGGGAGACACAGCAGCTCCGGGCTGGAGAGAACCGAGGGCGGAT 2145
Qy 701 AlaSerMetAlaCysArgIleValAlaAlaHisProLeuLeuLeuLeuHisIleu 720
Db 2146 GCCACAGAGGCTGCCGAGAGCTGGCGGTGGCGACCCGCTGCTCTCTCAGGCACTGTG 2205
Qy 721 PrometIleAlaAlaLeuLeuHisGlyArgTrpHisIleuAspPheGlnGlnPheArgGln 740
Db 2206 CCCATGATCGGGCGCTCTCTGACGCGCGACCCACCTCAACTTCAGAGAGTTCGGGCGAG 2265
Qy 741 GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuProHis 760
Db 2266 CAGAACCACTGAGGCTGCTCTCTGACAGTGGCGGCTGCTGAGGAGCTCTGACGCGCAC 2325
Qy 761 ValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgLeu 780
|||||

```

```

Db 2326 GTTTCGACAGCAGACACAGAGGGGCGCTGTGGAGCTGCTCTCTTCACTCCGCGTG 2385
Qy 781 LeuLeuAsnTrpArgIleSerSerArgHisLeuAlaAlaPheIleAsnIlePheValGln 800
Db 2386 CTGCTGAATTTACAGAGAGCTCCCGCAGCTGGGCTGTCTTATCAACAAGTTTGTGAG 2445
Qy 801 PheIleHisIleTyrlleThrTrpAsnAlaProAlaAlaIleSerPheLeuGlnHis 820
Db 2446 TTTCATCCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2505
Qy 821 AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuIleSerLeu 840
Db 2506 GCCGAGCCGCTCCAGAGCTGCTCTTGCACACAGTGAAGCTGTGATGCTGAAATCCCTC 2565
Qy 841 LeuAlaGlyLeuSerLeuProSerArgAspAspArgTrpAspArgGlyLeuAspGlnGln 860
Db 2566 CTTCGAGGGCTAGGCTGCGCCAGCAGAGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 2625
Qy 861 GlyGlnGlnGlnSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro 880
Db 2626 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2685
Qy 881 LeuThrAlaAlaGlnMetAlaProTrpMetIleValArgLeuSerArgGlyGlnThrValGln 900
Db 2686 CTGACCGCGCGCGAGATGGCCCTTACATGAAGCGCTTTCCCGGGGCCAAACGGGTGAG 2745
Qy 901 AspLeuLeuGlnValLeuSerAspIleAspGlnMetSerArgArgProGlnLeuLeu 920
Db 2746 GATCTGCTGGAGGTTCTGTGTGATACAGAGATGTCCCGGGGAGAGAGAGAGATCTCG 2805
Qy 921 SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGlnGlyCysCysArgAsn 940
Db 2806 ACCTCTCTCTGACCACTGACAGCGGCTGATGAGAGTGGCGGAGAGAGTGGCGGCAAC 2865
Qy 941 LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaPhe 960
Db 2866 CTGGCTCTTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2925
Qy 961 LeuProThrPheMetIleCysLeuGlnArgLeuGlnAspPheGlnValValGlnThrAlaLeu 980
Db 2926 CTGCCACGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2985
Qy 981 ArgAsnLeuProGlnTrpAlaLeuLeuLeuCysGlnGlnHisAlaAlaValLeuLeuHisArg 1000
Db 2986 CGSAGACTGCTGAGTACGCTCTCTGCTCAAGAGACAGCGGCTGTGCTCTCCACCGG 3045
Qy 1001 AlaPheLeuValGlyMetIleGlyGlnMetAspProSerAlaGlnIleSerValAlaLeu 1020
Db 3046 GCTTCTCTGGTGGGATGTACGGCCAGATGAGACCCAGCGCGAGATCTCCAGGCGCTGG 3105
Qy 1021 ArgIleLeuHisMetGlnAlaValMet 1029
Db 3106 AGGATCTCTCATATGAGGCGCTGATG 3132

```

RESULT 11

US-10-033-301-21

Sequence 21, Application US/10033301

Patent No. US20020098506A1

GENERAL INFORMATION:

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Ferrara, Napoleone

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Goddard, Audrey

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tamas, Daniel

APPLICANT: Watanabe, Collin K.

APPLICANT: Wood, William I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic



```

: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: P2930R1C6
: CURRENT APPLICATION NUMBER: US/10/033,301
: CURRENT FILING DATE: 2001-12-27
: PRIOR APPLICATION NUMBER: 60/095,325
: PRIOR FILING DATE: 1998-08-04
: PRIOR APPLICATION NUMBER: 60/112,851
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,145
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,511
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: 60/115,558
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/115,565
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/115,733
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/119,341
: PRIOR FILING DATE: 1999-02-09
: PRIOR APPLICATION NUMBER: 60/119,537
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: 60/119,965
: PRIOR FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: 60/162,506
: PRIOR FILING DATE: 1999-10-29
: PRIOR APPLICATION NUMBER: 60/170,262
: PRIOR FILING DATE: 1999-12-09
: PRIOR APPLICATION NUMBER: 60/187,202
: PRIOR FILING DATE: 2000-03-03
: PRIOR APPLICATION NUMBER: PCT/US99/12252
: PRIOR FILING DATE: 1999-06-02
: PRIOR APPLICATION NUMBER: PCT/US99/28634
: PRIOR FILING DATE: 1999-12-01
: PRIOR APPLICATION NUMBER: PCT/US99/28551
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US00/03565
: PRIOR FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: PCT/US00/05841
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: PCT/US00/08439
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: PCT/US00/14941
: PRIOR FILING DATE: 2000-05-30
: PRIOR APPLICATION NUMBER: PCT/US00/15264
: PRIOR FILING DATE: 2000-06-02
: PRIOR APPLICATION NUMBER: PCT/US00/32678
: PRIOR FILING DATE: 2000-12-01
: NUMBER OF SEQ ID NOS: 38
: SEQ ID NO 21
: LENGTH: 3437
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-033-301-21

Alignment Scores:
: Pred. No.: 0
: Score: 5241.00
: Best Local Similarity: 100.00%
: Query Match: 100.00%
: DB: 12

Length: 3437
Matches: 1029
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-929-769-7 (1-1029) x US-10-033-301-21 (1-3437)
QY 1 MeHhIleuValaIhSaIaMeTvaIleuLeuThreuglyProProArgaIa 20
Db 46 ATGCACATCTCTGCTGCATGCCATGCTGATCTCTGCGCGCCGCTCGAGCC 105
QY 21 AspAsSerGluPheGlnAlaLeuAspIleTirPheProGluGluysProleuPro 40
|||||

```

```

Db 106 GACGACGAGATTCCAGGCGCTGCGACATCTGCTTCGAGAGAGAACCATGCCCC 165
QY 41 ThhAlaPheLeuValaSprHrSerGluGluAlaLeuLeuProAspTirPheLeuysleu 60
Db 166 ACCGCTTCTGCTGCTGCACATCGAGAGAGCGCTGCTGCTTCTGCTGCTGAGAGCTG 225
QY 61 ArgMetIleArgSerGluValaLeuArgLeuValaSpAlaAlaLeuGlnAspLeuGluPro 80
Db 226 CGCATGATCGCTTCTGAGGTGCTCCGCTGTGTGAGCGCCCTGACAGGACCTGAGAGCGC 285
QY 81 GlnGlnLeuLeuLeuPheValaGlnSerPheGlyIleProValSerSerMetSerIysleu 100
Db 286 CACGACCTGCTGCTGTTCGTCAGTGTGGATCCGCTGTCACGATGACCAATCTC 345
QY 101 LeuGlnPheLeuAspGlnAlaValaIhSaIaAspProGlnThreugluGlnAsnIleMet 120
Db 346 CTCACATTCCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 405
QY 121 AspLysAsnTyrMetAlaIhSaIaLeuValaGlnIhSaIaGluArgGlyIleAspGlyIle 140
Db 406 GACAGAAATTACATGGCCCTGCTGAGGTCCACATGAGCGCGCGCTCCGAGAGG 465
QY 141 GlnThrPheIhSerLeuLeuThraIhSaIaSerLeuProProArgAspSerThraIaIa 160
Db 466 CAGACTTTCACCTCTTCTGCTCAGAGCTCCCTCCGCGCGCGAGACACAGAGCA 525
QY 161 ProLysProLysSerSerProGluGlnProIleGlyGlnIleArgIleArgValaGlyThr 180
Db 526 CCCAAACCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 585
QY 181 GlnLeuArgValaLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200
Db 586 CACCTCCGGGTGCTGGGCTGAGAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 645
QY 201 LeuSerProAspProArgTirPginSerSerSerProArgProValaIleAlaLeuGln 220
Db 646 CTCAGCGCGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 705
QY 221 GlnAlaLeuGlyGlnLeuAlaArgValaIleGlnIleSerProGluValaProGlyIle 240
Db 706 CAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765
QY 241 ThhValArgAlleuGlnAlaLeuAlaIhSaIaThrLeuLeuSerSerProIhSaIaGlyVala 260
Db 766 ACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
QY 261 ValMetSerMetIhSaIaArgSerIhSaIaPheLeuAlaCysProLeuLeuArgGlnLeu 280
Db 826 GTGATGTCATGACAGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885
QY 281 TyrGlnArgCysValaProGlnAspThrGlyPheSerSerLeuPheLeuysValleuLeu 300
Db 886 TACAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 945
QY 301 GlnMetLeuGlnIhPheLeuAspSerProGlnValaGlnIleGlyProLeuArgAlaGlnLeu 320
Db 946 CAGATGCTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005
QY 321 ArgMetLeuAlaSerGlnAlaIhSaIaSerAlaGlyArgArgLeuSerAspValaArgGlyIle 340
Db 1006 AGGATGCTTCCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1065
QY 341 LeuArgLeuAlaGluAlaLeuAlaIhSaIaPheArgGlnAspLeuGluValaIleSerSer 360
Db 1066 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1125
QY 361 ArgAlaValaIleAlaIhSaIaPheArgSerGlyGlnCysSerValaGluProAspLeuIle 380
Db 1126 CGTGCCTGATCCACACCTGAGGTCTGGGAGACAGTGCAGGCTGAGGAGGAGGAGGAGG 1185
QY 381 SerLysValaLeuGlnIleLeuIleGluValaIhSaIaSerProIhSaIaGlnIleLeuThr 400
Db 1186 AGCAAGTCTCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1245

```



OY	401	AlaPhePheSerLarThrAlaAspAlaAlaSerProPheProAlaCysLysProValVal	420
Db	1246	GCATTCCTTCCTGCCACTGCGGAGTGTGCTCCCGCTTCCAGCTGTAACCCGCTGTG	1305
OY	421	ValValSerSerLeuLeuLeuGlnGlnGlnGlnProLeuAlaGlyLysProGlyAla	440
Db	1306	GTGGTGAAGCTCCCTGCTGCTGCAGAGAGAGAGACCCTGGCTGGGGGAGACCGGGTGC	1365
OY	441	AspGlyLysSerLeuGlnAlaValAlaGlyLysProSerSerGlyLeuLeuValAspTrp	460
Db	1366	GAGCGTGCGACGCTGGAGGGCGGTGGGGCTGCTGTCAGGGCTCTTGTGTGACTGG	1425
OY	461	LeuGluMetLeuAspProGluValValSerSerCysProAspLeuGlnLeuArgLeuLeu	480
Db	1426	CTGGAATATGCTGCACCCCGAGGGTGTACAGCTGCGCCGACACTGCACACTCAGCTGCTC	1485
OY	481	PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProTrpLeuLeu	500
Db	1486	TTTCTCCGGAGGAAGGGCAAAAGGTCTCAGGGCCAGGTGCTCTCTTCCTGCTTCACTCTG	1545
OY	501	ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysILaArgValLeuLeu	520
Db	1546	ACCTCTTTCACGCTACGTCCACTGAGCCACACTGACACTGCATCCGACTCTGCTG	1605
OY	521	GlyLysSerArgLysGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys	540
Db	1606	GGCAAGAGCGGGAAACAGAGGTTTCAGACCCTGCTGCTCTGACTTCCTGGGGCTGC	1665
OY	541	ILeHisValProArgLysIleTrpGlnGlyAlaArgAspGlnArgThrProGlnLysArgArgLys	560
Db	1666	ATTCATGTCTCTGCACCTGTGGCAGGGGGGGGAGCACGGCACCCCGCAAGAGGGGGAG	1725
OY	561	GluLeuValLeuArgValGlnGlyProGluLeuILeSerLeuValGlnLeuLysLeuAla	580
Db	1726	GAGCTGGGTGCTGGGGTCCAGGGGCCGGAGACTATACGCTGTGTGACTGATCTGTGCC	1785
OY	581	GluAlaGluThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuILeGlnAlaArg	600
Db	1786	GAGCGGGAGACGCGAGCCAGGAGCGGGGAGACAGCCGCTGCAGCTCATCCAGGCCGG	1845
OY	601	LeuProLeuLeuLeuSerCysCysGlyAspAspGlnSerValArgLysValThrGlu	620
Db	1846	CTGCCCCCTGCTCACTCTCTCTGTGGGAGAGTGTGTGAGGAAGTGTGAGGAG	1905
OY	621	HisLeuSerGlyCysILeGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArgAsp	640
Db	1906	CACCTGTACAGGCTCATTCACGAGGTGGGGAGACAGCTGCTGGGAAGGGCTCCGAGAC	1965
OY	641	LeuLeuLeuGlnLeuTrpLeuGlnArgProGluLeuArgValProValProGluValLeu	660
Db	1966	CTTCTCTGCAGCTCATCTACAGCGGCGGAGCTGCGGGTGCCTGTGAGTGTCTTA	2025
OY	661	LeuHisSerGlnGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuILeHisArg	680
Db	2026	CTGCACACCGAAGGGCTGCCACAGCAGCGTGTGCACACTGCAGCGACTATCCACGCC	2085
OY	681	PheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAlaAsp	700
Db	2086	TTTCTATCAGCTCTTGGCGGACACACAGCACTCCCGGGCTTGGAAACCGAGGGCGGAT	2145
OY	701	AlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuLeuAlaHisLeu	720
Db	2146	GCCAGCATGGCTTCCCGAAGAGTGGGGTGGGACCCGCTCTCTCTCTGAGGCACTG	2205
OY	721	ProMetIleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPheGlnGluPheArgGln	740
Db	2206	CCCATTGATCGGGGCTCTGCACAGGGCGGACCACTCAACTTCCAGAGATTCTGGGAG	2265
OY	741	GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuGlnProHis	760
Db	2266	CAGAACCACTGACCTCTTCTGACAGGTGTGGGCTCTGTGAACTCTGTGACGCGGAC	2325

OY	761	ValpHekArgSerGlnHisGlnGlnValAlaLeuTyrAspCysLeuLeuSerPheIleArgLeu	780
Db	2326	GTGTTCCGACAGCAGCACAGGGGGGCGCTGGGACTGGCTTCTCTTCATCCGGCTG	2385
OY	781	LeuLeuAsnTyrArgLysSerSerArgHisLeuAlaAlaPheIleAsnLysPheValGln	800
Db	2386	CTGTGATTAATACAGGAAGTCTCTCCGCCATCTGTGGCTTCATCATCAACAAGTTGTGAG	2445
OY	801	PheIleHisLysTyrIleThrTyrAsnAlaProAlaAlaIleSerPheLeuGlnHis	820
Db	2446	TTTCATCCATAAGTACATTACTTCATCAATGCCGCCAGCAGCCATCTCCTTCAGACAGAC	2505
OY	821	AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu	840
Db	2506	GCCACCCCGCTCCACGACGCTGTCTTCCTTCGACAAAGAGACTTGATCTGAAATCCCTC	2565
OY	841	LeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArgGlyLeuAspLuglu	860
Db	2566	CTTCGAGGGCTCAACCTCGCCACAGAGAGAGACAGACCGACGACGAGCCCTGGACAAAG	2625
OY	861	GlyGlnGluGluSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro	880
Db	2626	GGCAGAGAGGAGAGCTACGCCGCGCTCCTTGCCCTGGTCAGCGTCTCCCTGTTCACCCCT	2685
OY	881	LeuThrAlaAlaGluMetAlaProTyrMetLysArgLeuSerArgGlyGlnThrValGlu	900
Db	2686	CTGACCCCGCGCGGAGAGGGCCCTACATGAAACGGCTTTCCGGGGCCAAACGGTGGAG	2745
OY	901	AspLeuLeuGluValLeuSerAspIleAspGluMetSerArgArgProGluLeuLeu	920
Db	2746	GATCTGCTGGAGGTTCTGAGTGAATAGACAGAGATGTCCCGCGGAGACCGAAGTCTCTG	2805
OY	921	SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGluGluCysArgAsn	940
Db	2806	AGCTTCTTCTGACCAACCTGCAGCGCGCTGATGAGCTCGGCGAGAGAGTGGCCGCAAC	2865
OY	941	LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerTlleAlaAlaPhe	960
Db	2866	CTCCGCTTCACCGCCGCGCCCTCCATGCAAAACAGCCCGAGATTGCAACCCCTTTTC	2925
OY	961	LeuProThrPheMetTyrCysLeuGlySerGlnAspPheGlyValValGlnThrAlaLeu	980
Db	2926	CTGCCCACTTCATGTACTCTGCTGGCAGCCAGGACTTTGAGAGTGGTGCAGACGGCCTC	2985
OY	981	ArgAsnLeuProGluTyrAlaLeuLeuCysGlnGluHisAlaAlaValLeuLeuHisArg	1000
Db	2986	CGGAACCTGCCTGAGTACGGTCTCTGTCGCAAGACACAGCGGCTGTGCTCCACCGG	3045
OY	1001	AlaPheLeuValGlyMetTyrGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeu	1020
Db	3046	GCCCTCTCGTGGGAGCATGTCAGCGCCAGATGAGACCCAGCGCGAGATCTCCGAGGCCCTG	3105
OY	1021	ArgIleLeuHisMetGluAlaValMet	1029
Db	3106	AGGATCCTGCATATGAGAGCGCGTGAAG	3132
RESULT 12			
US-10-033-326-21			
; Sequence 21, Application us/10033326			
; Patent No. US2020098507A1			
; GENERAL INFORMATION:			
; APPLICANT: Botstein, David			
; APPLICANT: Desnoyers, Luc			
; APPLICANT: Ferrara, Napoleone			
; APPLICANT: Fong, Sherman			
; APPLICANT: Gao, Wei-Qiang			
; APPLICANT: Goddard, Audrey			
; APPLICANT: Gurney, Austin L.			
; APPLICANT: Pan, James			
; APPLICANT: Roy, Margaret Ann			
; APPLICANT: Stewart, Timothy A.			
; APPLICANT: Thomas, Daniel			
; APPLICANT: Watanabe, Colin K.			

APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2930R1C8  
CURRENT APPLICATION NUMBER: US/10/033,336  
CURRENT FILING DATE: 2001-12-27  
PRIOR APPLICATION NUMBER: 60/095,325  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/112,851  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,145  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,511  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/115,558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,565  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,733  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119,341  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: 60/119,537  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/119,965  
PRIOR FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: 60/162,506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 60/170,262  
PRIOR FILING DATE: 1999-12-09  
PRIOR APPLICATION NUMBER: 60/187,202  
PRIOR FILING DATE: 2000-03-03  
PRIOR APPLICATION NUMBER: PCT/US99/12252  
PRIOR FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: PCT/US99/28634  
PRIOR FILING DATE: 1999-12-01  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US00/03565  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: PCT/US00/05841  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: PCT/US00/08439  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: PCT/US00/14941  
PRIOR FILING DATE: 2000-05-30  
PRIOR APPLICATION NUMBER: PCT/US00/15264  
PRIOR FILING DATE: 2000-06-02  
PRIOR APPLICATION NUMBER: PCT/US00/32678  
PRIOR FILING DATE: 2000-12-01  
NUMBER OF SEQ ID NOS: 38  
SEQ ID NO 21  
LENGTH: 3437  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-033-326-21

Alignment Scores:  
Pred. No.: 0 Length: 3437  
Score: 5241.00 Matches: 1029  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 12 Gaps: 0

US-09-929-769-7 (1-1029) x US-10-033-326-21 (1-3437)

Qy 1 MethsIleuValValIhIsalameValIleleuLeuThrIleuGlyProPArgAla 20  
Db 46 ATGCACATCTCTGCTGCATGCCATGTCATCTCTGTCGCTGGCCGCTCGAGCC 105

Qy 21 AspAspSerGluPheGlnAlaLeuAspIleTrpPheProGluIlyProLeuPro 40  
Db 106 GAGCAGACGAGATGTCAGGCGCTGCTGACATCTGTTCCGAGAGACCATGCCCC 165  
Qy 41 ThrAlaPheLeuValAspThrSerGluGlnAlaLeuLeuProAspTrpLeu 60  
Db 166 ACCGCTCTCTGCTGACATCGAGAGAGCCGCTGCTCTCTGCTGCTGACACTG 225  
Qy 61 ArgMetIleArgSerGluValIleArgLeuValAspAlaAlaLeuGlnAspLeuGluPro 80  
Db 226 CGCATGATCCGTTCTGAGTGGCTCCGCTGCTGAGCCGCCCTGCAGAGACCTGAGCCG 285  
Qy 81 GlnGlnLeuLeuLeuPheValGlnSerPheGlyIleProValSerMetSerIlyLeu 100  
Db 286 CAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 345  
Qy 101 LeuGlnPheLeuAspGlnAlaValAlaIleAspProGlnThrLeuGlnIleMet 120  
Db 346 CTCACAGTCTGACACGAGGAGTGGCCACGACCCGACACTTGGAGCAGAACATCATG 405  
Qy 121 AspLysAsnTrpMetAlaIleValIleValIleGlnIleGlnIleValArgGlyAlaSerGly 140  
Db 406 GACAGAAATTACATGAGCCACCTGCTGAGTGCAGATGAGCGGCGCTCGGAGGC 465  
Qy 141 GlnThrPheIleSerLeuLeuThrAlaSerLeuProProArgArgAspSerThrGlnAla 160  
Db 466 CAGACTTCCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525  
Qy 161 ProLysProLysSerSerProGlnGlnProIleGlnIleValArgIleValGlyThr 180  
Db 526 CCCAAGCAATGAGCAGACCCAGAGCAGCCATAGCCAGGCGCGGANTCGGTTGGGAGCC 585  
Qy 181 GlnLeuArgValLeuGlyProGlnAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200  
Db 586 CACCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 645  
Qy 201 LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220  
Db 646 CTCAGCCCGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 705  
Qy 221 GlnAlaLeuGlyGlnIleValAlaValIleValIleValIleValIleValIleValIleVal 240  
Db 706 CAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765  
Qy 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProIleGlyAlaLeu 260  
Db 766 ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825  
Qy 261 ValMetSerMetIleAspSerIlePheLeuAlaCysProLeuLeuArgGlnLeuCysGln 280  
Db 826 GTATGCTCATGACCGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885  
Qy 281 TyrGlnArgCysValProGlnAspTrpGlyPheSerSerLeuPheLeuValLeuLeu 300  
Db 886 TACCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 945  
Qy 301 GlnMetLeuGlnTrpLeuAspSerProGlyValIleGlyIleProLeuArgAlaGlnLeu 320  
Db 946 CAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005  
Qy 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyIleLeu 340  
Db 1006 AGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1065  
Qy 341 LeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValIleSerSerThrVal 360  
Db 1066 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125  
Qy 361 ArgAlaValIleAlaThrLeuArgSerGlyGlnIleCysSerValGlnProAspLeuIle 380  
Db 1126 CGTGCCTATCCACACCTGAGAGTCTGGGAGACAGTGCAGTGCAGCCGAGCTGATC 1185  
Qy 381 SerIlyValLeuGlnIleGlnIleGlnIleValArgSerProIleLeuGlnIleLeuLeuThr 400

Db	1186	AGCAAGTCCTCCAGGGGGCTGATCGAGGTGAGGTGCTCCCGACCTGGAGAGCTGTGACT	1245
Qy	401	AlaPhePheSerAlaThrAlaAspAlaAlaSerProPheProAlaCysLysProValAla	420
Db	1246	GCAATCTTCCTCTGCACTGGCGAGTGTGCTCCCGCTTTCACAGCTGTAAACCCGTTGTG	1305
Qy	421	ValValSerSerLeuLeuLeuGlnGlnGlnGlnProLeuAlaGlyGlyLysProGlyVala	440
Db	1306	GTGGTGAGCTTCCTGCTGCTGCGACAGAGAGAGAGACCCCTTGCTGGGGGGAGACCGGATGCG	1365
Qy	441	AspGlyLysSerLeuGlnAlaValaLargLeuGlyProSerSerGlyLeuLeuValaAspTrp	460
Db	1366	GACGGTGGACAGCTGGAGGCCGTGGGGCTGGGCCCTGTGACGGCTCTGATGAGACTGG	1425
Qy	461	LeuGlnMetLeuAspProGlnValaValaSerSerCysProAspLeuGlnLeuAlaArgLeu	480
Db	1426	CTGGAAATGCTGGACCCCGAGGGTGGTACACACTGTGCCCCGACCTCAGCTCAGGCTGTG	1485
Qy	481	PheSerArgArgLysGlyLysGlyGlnAlaGlnValaProSerPheArgProTrpLeuLeu	500
Db	1486	TTTCTCCGGAGAGAGGGCAAAAGGTACAGGCCCAAGTGCCTCTTTCGCTCCTACTCTG	1545
Qy	501	ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeuLeu	520
Db	1546	ACCCCTTTCACAGCATCAGTCCAGCTGGCCCAACACTGCACAAATGCATCCGAGTCTGCTG	1605
Qy	521	GlyLysSerArgGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys	540
Db	1606	GGCAAGAGCCGGACAGAGAGGTTCAGCCCTTGCTCTCTTGACTTCTCTGGGGCTGCG	1665
Qy	541	IleHisValProArgIleTrpGlnGlyArgAspGlnArgTrpProGlnLysArgArgGlu	560
Db	1666	ATCCATGTCTCTCGCATCTGGACGGGGGGGAGACAGCCACCCCGCAAAAGCGGGGGAG	1725
Qy	561	GluLeuValLeuArgValaGlnGlyProGlnLeuIleSerLeuValGlnLeuIleLeuAla	580
Db	1726	GAGCTGGTGGCTGGGGGTCCAGGGCCCGAGACTATAGCTGTGTGAGCTGATCTGTGGCC	1785
Qy	581	GluAlaGlnThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAlaArg	600
Db	1786	GAGCGGAGACCGGAGACCAGAGCGGGGACACAGCGCGCTCAGCCCTCATCAGAGCCCGG	1845
Qy	601	LeuProLeuLeuLeuSerCysCysGlyAspAspGlnSerValaLargLysValaThrGlu	620
Db	1846	CTGGCCCTGGCTCAGCTCCTCTCTGTGGGAGAGAGAGTGTGAGAGAGGTGACGGAG	1905
Qy	621	HisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValaLeuGlyArgArgCysArgAsp	640
Db	1906	CACCTGTACGGCTGTACACCAAGAGTGGGGAGACAGCTGTGGAGAGCGCTCCCGAGAC	1965
Qy	641	LeuLeuLeuGlnLeuTrpLeuGlnArgProGlnLeuArgValaProValaProGlnValaLeu	660
Db	1966	CTTCTCTCTGACGCTCTACCTACAGCGCGGAGCTGGGGGTCCCGCTGCTGAGAGTCTTA	2025
Qy	661	LeuHisSerGlnGlyAlaAlaSerSerSerValaCysLysLeuAspGlyLeuIleHisArg	680
Db	2026	CTGCACACGAAAGGGCTGCCACACAGAGGTGTGACACTGTGAGAGACTATCTCACCGC	2085
Qy	681	PheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAlaAsp	700
Db	2086	TTTCATCACGCTCTTGGCGACACACAGCATCCCGGGCGTTGGAGAACCGAGGGCGGAT	2145
Qy	701	AlaSerMetAlaCysArgGlySerLeuAlaAlaHisProLeuLeuLeuAlaArgHisLeu	720
Db	2146	GGCCAGCATGGCTCTGCCGAAGCTGGGGGTGGCGCACCCGCTCTCTCCTCAGGCACTG	2205
Qy	721	ProMetIleAlaAlaLeuLeuHisGlyArgTrpHisLeuAspPheGlnGlyIlePheArgGln	740
Db	2206	CCCATGATCCGGGGCTCTCGACAGGGCCGACCCACTCAACTTCCAGAGAGTCCGGGAG	2265
Qy	741	GlnAsnHisLeuSerCysPheLeuHisValaLeuGlyLeuLeuGlnLeuLeuGlnProHis	760

Db	2266	CAGAACCAACCGTGAAGCTCTCCCTCCGACAGGCTGGGCGCTGCTGGAGCTGCTGACAGCCGAC	2325
Oy	761	ValPheargserGlnHisGlnGlyAlaLeuTTPaspCysLeuLeuSerPheIleargLeu	780
Db	2326	GTGTTTCGCGACGCGACCAAGGGGGGCGTGGGACGTGGCTTCTGCTTCATCCGCTC	2385
Oy	781	LeuLeuasnTrpArgIysSerSerArgHisIleuAlaAlaPheIleAsnIysPheValGln	800
Db	2386	CTGCTGAATTACAGAGAAGTCTCTCCGCGCATCTGGCTTCCCTTATCAACAAGTTTGTGAG	2445
Oy	801	PheIleHisIysTrpIleThrTrpAsnAlaProAlaAlaIleSerPheLeuGlnHis	820
Db	2446	TTTCATTCATTAAGTACATTACTTACCTACAAATGCCCCGACGAGCATCTCCCTCTCAGAACAC	2505
Oy	821	AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerIleu	840
Db	2506	GCCACCCCGCTCCACACGCGCTGCTTCCTTGACAAACAGATGACTGTGTATCCTGAATAACCTTC	2565
Oy	841	LeuAlaGlyLeuSerIleuProSerArgAspAspArgTrpHisAspArgIleuAspGlu	860
Db	2566	CTTCACAGGGCGTCAACCTCGCCACACAGGAAGACAGGACCGACGAGGCTTGACGACAAG	2625
Oy	861	GlyGluGluGlnSerIleArgIleGlySerLeuProLeuValSerValSerLeuPheThrPro	880
Db	2626	GGCAGAGAGGAGAGACTACGCGGCTCTTGGCCCTGCTGAGCGTCCCTGTCACCCCT	2685
Oy	881	LeuThrAlaAlaGluMetAlaProTrpMetLysArgLeuSerArgGlyGlnThrValGlu	900
Db	2686	CTGACCCCGCGCGAGATGGCCCCCTTACATGTAAGAAAGCGCTTCCCGGGGCGAAACGTGTGAG	2745
Oy	901	AspLeuLeuGlnValLeuSerAspIleAspGluMetSerArgArgArgProGluIleLeu	920
Db	2746	GATCTGCTGGAGGTTCTGAGTGTGACATAGACAGAGATGTCCCGCGGAGACCCGACATCTCTG	2805
Oy	921	SerPhePheSerTrpAsnLeuGlnArgLeuMetSerSerAlaGluGlnCysCysArgAsn	940
Db	2806	AGCTTCTTCTCGACCACTCTGCACGCGCTGATGAGCTCGGCGGAGAGAGTGTGGCGGCAC	2865
Oy	941	LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaAlaPhe	960
Db	2866	CTCGCCTTCACGCTCGGCCCTCGCCTCCATGACAAACAGCCGACGATTGACACCCCTTTC	2925
Oy	961	LeuProThrPheMetTrpCysLeuGlySerGlnAspPheGluValValGlnThrAlaLeu	980
Db	2926	CTTGCCACGTTTCATGATCTCTCGGGCAGCGACGACACTTGTGAAGTGTGTACAGCGCCTTC	2985
Oy	981	ArgAsnLeuProGluTrpAlaLeuLeuCysGlnGlnHisAlaAlaValLeuLeuHisArg	1000
Db	2986	CGGAACCTGGCTAGTACGTCTCTGCTGGCAAGACACAGCGGCTGTGCTGTCCACGG	3045
Oy	1001	AlaPheLeuValGlyMetTrpGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeu	1020
Db	3046	GCCTTCTCTGGGGGCATGTACGGCCAGATGGAGCCCGACGCGCAGATCTCGAGGCGCTG	3105
Oy	1021	ArgIleLeuHisMetGluAlaValMet 1029	
Db	3106	AGGATCTCGCATATGAGAGGCGCGTGATG 3132	
RESULT 13			
US-09-822-849A-276			
; Sequence 276, Application US/09822849A			
; Patent No. US20020045170A1			
GENERAL INFORMATION:			
; APPLICANT: Wong, Gordon G.			
; APPLICANT: Clark, Hilary			
; APPLICANT: Flechtel, Kim			
; APPLICANT: Agostino, Michael J.			
; APPLICANT: Howes, Steven H.			
; APPLICANT: Resnick, Richard J.			
; APPLICANT: Gulukota, Kamalakar			
; APPLICANT: Graham, James R.			
; APPLICANT: Genetics Institute, Inc.			
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS			

```

, RESULT 13
, US-09-822-049A-276
, Sequence 276, Application US/09822849A
, Patent No. US20020045170A1
, GENERAL INFORMATION:
, APPLICANT: Wong, Gordon G.
, APPLICANT: Clark, Hilary
, APPLICANT: Fechtel, Kim
, APPLICANT: Agostino, Michael J.
, APPLICANT: Howes, Steven H.
, APPLICANT: Resnick, Richard J.
, APPLICANT: Gujukota, Kamalakr
, APPLICANT: Graham, James R.
, APPLICANT: Genetics Institute, Inc.
, TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS

```

```

: FILE REFERENCE: GIN 6403
: CURRENT APPLICATION NUMBER: US/09/822,849A
: CURRENT FILING DATE: 2001-09-04
: PRIOR APPLICATION NUMBER: 60/195,582
: PRIOR FILING DATE: 2000-04-06
: NUMBER OF SEQ ID NOS: 598
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 276
: LENGTH: 2824
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-822-849A-276

Alignment Scores:
Pred. No.: 0          Length: 2824
Score: 4211.00       Matches: 829
Percent Similarity: 98.81%  Conservative: 1
Best Local Similarity: 98.69%  Mismatches: 0
Query Match: 80.35%  Indels: 10
DB: 10          Gaps: 1

US-09-929-769-7 (1-1029) x US-09-822-849A-276 (1-2824)

QY 200 ProLeuSerProaspProargTrpGlnSerSerProargProvalAlaLeuAlaLeu 219
Db 2 CGCCTACAGCCCGGACCTCGGTGGCAGAGCTCCAGCCCGCGGCTGCGCTGCGCTG 61
QY 220 GlnGlnAlaLeuGlnGlnGlnLeuAlaArgValValGlnGlnGlnProGlnValProGln 239
Db 62 CACAGAGCCCTGGGCGGAGAGCTGGCCGCGCTCCAGGGGACCCCGAGAGCTCGGGC 121
QY 240 IleThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerProHisGlyGlyAla 259
Db 122 ATACAGGTGGTGTCTCTGCGAGGCCCTGCGCACCTGCTCAAGCTCCACAGCGCGGTGC 181
QY 260 LeuValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuCys 279
Db 182 CTGGTGTGTGCATGACGACCTGAGCTTCTGCTGCGCGCTGCGCGCTGCGCGCTGCG 241
QY 280 GlnTyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValLeu 299
Db 242 CACTACAGAGCGGTGTGTGCGACAGAGACCGGCTTCTCTGCTGCTGCTGCTGCTGCTG 301
QY 300 LeuGlnMetLeuGlnTrpLeuAspSerProGlyValGlnGlnGlnProLeuArgAlaGln 319
Db 302 CTCAGATGTGTGACGTGGTGGACAGCCCTGCGGTGGAGGGCGGCGCTGCGGGCACAG 361
QY 320 LeuArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGly 339
Db 362 CTCAGAGATGTGTGCGACAGCCGCTGAGCGGCGGCGAGCTCACTGATGTGCGAGGGGG 421
QY 340 LeuLeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValAlaSerSerThr 359
Db 422 CTCCTGGGCTGGCGGAGCGGCTGCGCTTCTGCTGAGACCTGAGAGTGTGCTGCTGCTG 481
QY 360 ValArgAlaValIleAlaThrLeuArgSerGlyGlnGlnCysSerValGluProAspLeu 379
Db 482 GTCCGTCCTCTCATGCGCACCTGAGGTCTGGGGACAGTGCAGCTGAGAGCCGACCTG 541
QY 380 IleSerTrpValLeuGlnGlnGlnLeuIleGlyValArgSerProHisGlnGlnGlnLeu 399
Db 542 ATACAGCAAGCTCTCCAGGGGCTGATGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 601
QY 400 ThrAlaPhePheSerAlaThrAlaAspAlaAlaSerProPheProAlaCysLysProVal 419
Db 602 ACGCATTTCTTCTCTGCACTGGGAGTGTGCTCCCGCTTCCAGCTGTAAGCCGCTT 661
QY 420 ValValValSerSerLeuLeuGlnGlnGlnGlnProLeuAlaGlyGlyLysProGly 439
Db 662 GTGTGTGTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721
QY 440 AlaAspGlyLysLeuGlnGlnAlaValArgLeuGlyProSerSerGlyLeuLeuValAsp 459

```

```

Db 722 GCGAGCGGTGGCAGCCCTGAGGCCGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 781
QY 460 TrpLeuGlnMetLeuAspProGlnValValSerSerCysProAspLeuGlnLeuArgLeu 479
Db 782 TGGCTGGAATGCTGGAGCCCGGAGGTGTGACAGAGCTGCCCGACCTGACAGCTGAGCTG 841
QY 480 LeuPheSerArgArgGlyGlyGlyGlnAlaGlnValProSerPheArgProTrpLeu 499
Db 842 CTCCTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 901
QY 500 LeuThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeu 519
Db 902 CTGACCTCTTCAAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
QY 520 LeuGlyLysSerArgGlnGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAla 539
Db 962 CTGGGCAAGAGCGGGAACAGAGGTGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1021
QY 540 CysIleHisValProArgIleTrpGlnGlyArgAspGlnArgThrProGlnLysArgArg 559
Db 1022 TGCATCCATGTTCTCTGCAATCTGCGACGGGCGGAGCCAGCCCGCAGAGAGCGGCGG 1081
QY 560 GlnGlnLeuValLeuArgValGlnGlnGlnProGlnLeuLeuIleSerLeuValGlnLeu 579
Db 1082 GAGAGCTGTGTGTGCGGTGTCAGGGCCCGAGCTCATGAGCTGTGTGAGCTGATCTCTG 1141
QY 580 AlaGlnAlaGlnThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAla 599
Db 1142 GCGAGAGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
QY 600 ArgLeuProLeuLeuLeuSerCysCysGlyAspAspGlnSerValArgLysValThr 619
Db 1202 CGCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1261
QY 620 GlnHisLeuSerGlyCysIleGlnGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArg 639
Db 1262 GAGCAGCTGTGAGCTCATTCAGCACTGGGAGAGACAGCTGCTGGCGAGCGCTGCCGA 1321
QY 640 AspLeuLeuGlnGlnLeuTyrLeuGlnArgProGlnLeuArgValProValProGlnVal 659
Db 1322 GACCTTCTCGTGCAGCTCTACCTACAGCGGCGGAGCTGCGGGTGGCGCTGAGAGCTC 1381
QY 660 LeuLeuHisSerGlnGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuIleHis 679
Db 1382 CTACTGCACAGCGAAGGGGTGCGCACAGAGCTGTCAGAGCTGAGCGAGCTCATCTAC 1441
QY 680 ArgPheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnLysArgGlyAla 699
Db 1442 CGCTTCATCAGCTCTTCTGCGAGACCAAGCACTCCCGGGCTTGGAGAACCGAGGGGG 1501
QY 700 AspAlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuAlaArgHis 719
Db 1502 GATGCCAGCATAGGCTCGCGGAAGCTGGCGGTGGCGGACCGCTGCTGCTGCTGCTGCTG 1561
QY 720 LeuProMetIleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPheGlnGlnPheArg 739
Db 1562 CTGCCCATGATCGGGGGCTCTGTCAGGGCGGACCCACCTCACTTCAGAGAGTTCGG 1621
QY 740 GlnGlnAsnHisLeuSerCysPheLeuHisValLeuGlnGlnGlnGlnGlnGlnGlnPro 759
Db 1622 CAGCAGAACCACTGAGAGCTGCTTCTGACAGCTGTGGGCTGCTGCTGCTGCTGCTGCTG 1681
QY 760 HisValPheArgSerGlnHisGlnGlnValAlaLeuTrpAspCysLeuLeuSerPheLearg 779
Db 1682 CAGCTGTTCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1741
QY 780 LeuLeuLeuAsnTyrArgLysSerSerArgHisLeuAlaAlaPheIleAsnLysPheVal 799
Db 1742 CTGCTGTGAATTAACAGGAAGTCTCCCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1801
QY 800 GlnPheIleHisLysTyrIleThrTyrAsnAlaProAlaAlaIleSerPheLeuGlnLys 819
Db 1802 CAGTTTCATCAATAGTACATTACCTACATGCCCCAGCAGCAGCATCTCTCTGCTGAGAG 1861

```



QY 906 LeuSerAspIleAspGluMetSerArgArgProGluIleLeuSerPhePheSerThr 925  
 Db 841 CTGAGTACATTAGACGAGTGTCCCGGAGACCCGAGATCTGAGCTTCTCTGACC 900  
 QY 926 AsnLeuGlnArgLeuMetSerSerAlaGluGluCysCysArgAsnLeuAlaPheSerLeu 945  
 Db 901 AACCTGACGGGCTGATGAGCTGCGCGAGAGATGTTGCGCAACCTCGCTTCAGCTTG 960  
 QY 946 AlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaAlaPheLeuProThrPheMet 965  
 Db 961 GCCCTCGCTCCATGACGAACACCCGAGATGTCAGACCCCTTCCGCGCCAGCTTCATG 1020  
 QY 966 TyrCysLeuGlySerGlnAspPheGluValGlnThrAlaLeuArgAsnLeuProGlu 965  
 Db 1021 TACTGCTGGGACGACGAGCTTGTGAGTGTGTCAGACGCGCTCCGGAACCTGCTGAG 1080  
 QY 986 TyrAlaLeuLeuGlyGlnGlnHisAlaAlaValLeuLeuHisArgAlaPheLeuValGly 1005  
 Db 1081 TACGCTCTCTGTGCGACAGACGCGGCTGTCTGCTCCACCGGCTTCTGCTGAGG 1140  
 QY 1006 MetTyrGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeuArgIleLeuHisMet 1025  
 Db 1141 ATGTAGCGCCAGATGACCCGACGCGGAGATCTCCGAGGCGCTGAGATCTGATATG 1200  
 QY 1026 GluAlaValMet 1029  
 Db 1201 GAGCGCGTGATG 1212

## RESULT 15

US-10-040-739-1131/C

Sequence 1131. Application US/10040739

Patent No. US2002017635A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John

Lavalley, Edward

Racie, Lisa

Merberg, David

Treachy, Maurice

Spaulding, Vilki

TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS

NUMBER OF SEQUENCES: 1519

CORRESPONDENCE ADDRESSES:

ADDRESSER: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/040,739

FILING DATE: 07-Jan-2002

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/036,520

FILING DATE: 03-JUN-1998

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 1131:

SEQUENCE CHARACTERISTICS:

LENGTH: 639 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1131:  
 US-10-040-739-1131

## Alignment Scores:

Pred. No.:	2,62e-90	Length:	639
Score:	924.00	Matches:	188
Percent Similarity:	98.44%	Conservative:	1
Best Local Similarity:	97.92%	Mismatches:	2
Query Match:	17.63%	Indels:	1
DB:	9	Gaps:	0

US-09-929-769-7 (1-1029) x US-10-040-739-1131 (1-639)

QY 839 SerLeuLeuAlaGlyLeuSerLeuProSerArgAspArgThrAspArgGlyLeuAsp 858  
 Db 637 TCCTCTCTGACGAGGCTTCAAGCTTCCAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAG 578  
 QY 859 GluGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 878  
 Db 577 GAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 518  
 QY 879 ThrProLeuThrAlaAlaGluMetAlaProTyrMetTyrArgLeuSerArgGlyGlnThr 898  
 Db 517 ACCCTCTGACCGCGCGGAGATGCGCCCTACATGAACGAGCTTCCGCGGCGCAACG 458  
 QY 899 ValGluAspLeuGluGluValLeuSerAspIleAspGluMetSerArgArgArgProGlu 918  
 Db 457 GTGAGAGATGCTGCTGAGGAGTTCGAGATGACATGACAGAGATGTCGCGGAGAGACCG 398  
 QY 919 IleLeuSerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGluGlyGlyGly 938  
 Db 397 ATCTGAGCTTCTTCTGACCAACCTGACGCGGCTGATGAGTGGCGCGGAGAGTGTTCG 338  
 QY 939 ArgAsnLeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAla 958  
 Db 337 CGCAACCTGCGCTTCAAGCTTGGGCGCTGCTCCATGACAGACAGAGAGAGAGAGAGAG 278  
 QY 959 AlaPheLeu-ProThrPheMetTyrCysLeuGlySerGlnAspPheGluValGlnThr 978  
 Db 277 GCTTCTGCGCCNACGTTGATGATGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 218  
 QY 978 ValLeuArgAspLeuProGluTyrAlaLeuGlyGlnGlnHisAlaAlaValLeu 998  
 Db 217 GGCCTTCGGAACCTGCTGATGAGCTTCTGCTGCTCCAGAGACAGAGAGAGAGAGAGAG 158  
 QY 998 ValSerAlaPheLeuValGlyMetTyrGlyGlnMetAspProSerAlaGlnIleSerGly 1018  
 Db 157 CCACGCGGCGCTTCTGCTGAGGATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 98  
 QY 1018 ValLeuArgGlyLeuHisMetGluAlaValMet 1029  
 Db 97 GGCCTGAGAGATCTCATATGAGGCGCGGAG 64

Search completed: May 14, 2003, 02:38:44  
 Job time : 442 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:41:38 ; Search time 44 Seconds  
(without alignments)  
3116.247 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 1029  
Sequence: 1 MHILVHAMVILLTLGPRA.....MDPSAQISEALRIILMEAVM 1029

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size: 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/genescp-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/genescp-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/genescp-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/genescp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1029	100.0	1029	21 AAB24027	Human PRO3434 prot
2	1029	100.0	1029	21 AAY96736	PRO3434, a novel s
3	1029	100.0	1029	23 AAU77765	Tumour associated
4	504	49.0	650	21 AAB42495	Human ORFX ORF2259
5	404	39.3	404	21 AAY87196	Human secreted pro
6	404	39.3	404	22 AAE06173	Human gene 46 enco
7	404	39.3	404	23 ABG33997	Human secreted pro
8	327	31.8	327	22 AAE06086	Human gene 46 enco
9	327	31.8	327	23 ABG33908	Human secreted pro
10	327	31.8	328	21 AAY87109	Human secreted pro

11	43	4.2	161	22 AAM95497	Human reproductive
12	8	0.8	55	22 AAU52278	Protonibacterium
13	8	0.8	84	22 AAU64883	Protonibacterium
14	8	0.8	101	23 ABP34288	Human kinase-like
15	8	0.8	207	22 AAB58449	Drosophila melanog
16	8	0.8	209	23 AAU83137	Novel secreted pro
17	8	0.8	217	22 AAC81704	S. epidermidis ope
18	8	0.8	224	20 AAU72977	Rhodobacter sphaer
19	8	0.8	224	20 AAU76766	R. sphaeroides Adh
20	8	0.8	226	22 AAB82495	Formaldehyde trans
21	8	0.8	253	20 AAY08016	Mouse PA-I protein
22	8	0.8	311	22 AAB59518	Drosophila melanog
23	8	0.8	390	20 AAY05302	S. aureus protein
24	8	0.8	396	23 AAB39735	Staphylococcus epi
25	8	0.8	705	22 AAB94141	Human protein sequ
26	8	0.8	853	22 AAB95413	Human protein sequ
27	8	0.8	893	22 AAB93648	Human protein sequ
28	8	0.8	1137	22 AAB94750	Human protein sequ
29	8	0.8	1846	22 AAU00295	Interferon induced
30	8	0.8	2042	22 AAB59689	Drosophila melanog
31	7	0.7	27	18 AAM09063	Epsitein-Barr virus
32	7	0.7	27	18 AAM18617	Peptide #5051 enco
33	7	0.7	27	22 AAM1077	Peptide #5114 enco
34	7	0.7	27	23 AAG40522	Human peptide enco
35	7	0.7	27	23 AAU10698	cPLA2-adaptor pept
36	7	0.7	36	9 AAB80687	Peptide 88 from th
37	7	0.7	38	22 AAB50760	Human secreted pro
38	7	0.7	39	21 AAB63080	Human secreted pro
39	7	0.7	43	21 AAG55810	Arabidopsis thalia
40	7	0.7	46	21 AAG59085	Arabidopsis thalia
41	7	0.7	47	21 AAB34147	Gene 4, human secre
42	7	0.7	47	21 AAB25068	Plant SP encoded
43	7	0.7	48	22 AAB60198	Human secreted alp
44	7	0.7	48	22 AAB60198	Novel central nerv
45	7	0.7	50	23 AAB00236	Human ORFX protein

ALIGNMENTS

RESULT 1  
AAB24027 standard: Protein: 1029 AA.  
ID AAB24027;  
AC AAB24027;  
XX 25-JAN-2001 (first entry)  
XX DT  
XX DE Human PRO3434 protein sequence SEQ ID NO:23.  
XX DE Human: tumour; diagnosis: neoplastic disease; proliferation; cancer;  
XX KW Identification; tumorigenesis; anticancer; detection.  
XX OS Homo sapiens.  
XX PN WO200053750-A1.  
XX PD 14-SEP-2000.  
XX PF 02-DEC-1999; 99WO-US28551.  
XX PF 01-SEP-1999; 99WO-US05028.  
XX PR 01-SEP-1999; 99WO-US20111.  
XX PR 29-OCT-1999; 99US-0162506.  
XX PR 30-NOV-1999; 99WO-US28313.  
XX PR 01-DEC-1999; 99WO-US28634.  
XX PA (GETH ) GENENTECH INC.  
XX PI Botstein D, Goddard A, Gurney AL, Roy MA, Matanabe CX, Wood WT.  
XX WPI: 2000-594320/56.  
XX DR N-PSDB: AAC58109.



XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit  
 PT the growth of tumors in mammals, and to identify inhibitors of PRO  
 PT polypeptide activity or expression -  
 XX  
 PS Claim 61: Fig 14; 226pp; English.  
 XX  
 CC The present invention describes an antibody that binds to a human  
 CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;  
 CC PRO3434; PRO1927; PRO3567; PRO1293; PRO1344; PRO4344; PRO4354;  
 CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has  
 CC anticancer activity and can be used to diagnose tumors in mammals, by  
 CC detecting complex formation when the antibody is contacted with test  
 CC cells. Increased expression of genes encoding (I) can also be detected  
 CC to diagnose tumors. Agents which inhibit the activity of (I),  
 CC especially the antibodies, or an antisense oligonucleotide which  
 CC hybridises to genes encoding (I), can be used to inhibit tumour growth,  
 CC preferably by inducing cell death. Methods from the present invention  
 CC can be used to identify compounds which inhibit the biological activity  
 CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation  
 CC probes used in examples from the present invention for human PRO  
 CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human  
 CC PRO polynucleotide and protein sequences given in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 1029 AA:  
 Query Match 100.0%; Score 1029; DB 21; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 601 LPLLSCCGDDSVKRVTEHLSGCCIOQWGSVLRRCRDLLOLYQRELPVPEVL 660  
 QY LHSEGAASSVCKIDGILHREFITLADTSDSRALENAGDASMACRLAVAHPLLLRL 720  
 Db 661 LHSEGAASSVCKIDGILHREFITLADTSDSRALENAGDASMACRLAVAHPLLLRL 720  
 QY 721 PMIAALLHGRTHLNFQEFROQNLSCFLHVLGLLELLQPHVSEHQALMDCLLSTRL 780  
 Db 721 PMIAALLHGRTHLNFQEFROQNLSCFLHVLGLLELLQPHVSEHQALMDCLLSTRL 780  
 QY 781 LWNKRSRLAFLPINKVQFIHKYTYNAPPAISFLQKADPLHDSPNSDLVYKSL 840  
 Db 781 LWNKRSRLAFLPINKVQFIHKYTYNAPPAISFLQKADPLHDSPNSDLVYKSL 840  
 QY 841 LAGSLPSRDRDRDRCDEGESESSAGSLPVSVSLFTPLTAENAPYKRLSRGQTV 900  
 Db 841 LAGSLPSRDRDRDRCDEGESESSAGSLPVSVSLFTPLTAENAPYKRLSRGQTV 900  
 QY 901 DLLEVLSDIDEMSRREILSFESTNLQRLMSSAECCRNLAESLALRSMQNSPSIAAF 960  
 Db 901 DLLEVLSDIDEMSRREILSFESTNLQRLMSSAECCRNLAESLALRSMQNSPSIAAF 960  
 QY 961 LPFTMYCLGSQDEEVQVOTALRNPETVALLQEHAAVILHRAFLVGMVGMDPSAQISEAL 1020  
 Db 961 LPFTMYCLGSQDEEVQVOTALRNPETVALLQEHAAVILHRAFLVGMVGMDPSAQISEAL 1020  
 QY 1021 RILHMEAVM 1029  
 Db 1021 RILHMEAVM 1029

RESULT 2  
 AAY96736  
 ID AAY96736 standard; Protein: 1029 AA.  
 XX AAY96736:  
 AC 26-SEP-2000 (first entry)  
 DT  
 XX  
 DE PRO3434, a novel secreted protein.  
 XX  
 KW PRO3434; secreted protein; transmembrane protein; recombinant production;  
 KM gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Peptide 1..16  
 FT /label= Signal\_peptide  
 FT Modified-site 91..97  
 FT /note= "N-myristoylation site"  
 FT Modified-site 136..142  
 FT /note= "N-myristoylation site"  
 FT Modified-site 154..158  
 FT /note= "CAMP- and cGMP-dependent protein kinase  
 FT phosphorylation site"  
 FT Modified-site 224..230  
 FT /note= "N-myristoylation site"  
 FT Modified-site 329..333  
 FT /note= "amidation site"  
 FT Modified-site 331..335  
 FT /note= "CAMP- and cGMP-dependent protein kinase  
 FT phosphorylation site"  
 FT Modified-site 435..441  
 FT /note= "N-myristoylation site"  
 FT Modified-site 439..445  
 FT /note= "N-myristoylation site"  
 FT Modified-site 443..449  
 FT /note= "N-myristoylation site"  
 FT Modified-site 616..620  
 FT /note= "CAMP- and cGMP-dependent protein kinase  
 FT phosphorylation site"  
 FT Modified-site 634..638



```

FT      /note= "amidation site"
FT      665..671
FT      Modified-site
FT      /note= "N-myristoylation site"
FT      698..704
FT      Modified-site
FT      /note= "N-myristoylation site"
FT      783..789
FT      /note= "cAMP- and cGMP-dependent protein kinase
FT      phosphorylation site"
FT      891..895
FT      /note= "cAMP- and cGMP-dependent protein kinase
FT      phosphorylation site"
FT      WO200036102-A2.
XX      22-JUN-2000.
XX      01-DEC-1999; 99WO-US28634.
XX      16-DEC-1998; 98US-0112851.
XX      16-DEC-1998; 98US-0113145.
XX      22-DEC-1998; 98US-0113511.
XX      12-JAN-1999; 99US-0115558.
XX      12-JAN-1999; 99US-0115565.
XX      12-JAN-1999; 99US-0115733.
XX      09-FEB-1999; 99US-0119341.
XX      10-FEB-1999; 99US-0119537.
XX      12-FEB-1999; 99US-0119665.
XX      02-JUN-1999; 99WO-US12252.
XX      (GETH ) GENENTECH INC.
XX      Botstein D, Desnovers L, Ferrara N, Fong S, Gao W, Goddard A;
XX      Guney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
XX      Wood WT;
XX      WPI: 2000-431586/37.
XX      N-PSDB: AAA51266.
XX      Isolated nucleic acid molecule encodes a PRO polypeptide which is a
XX      transmembrane polypeptide
XX      Claim 12; Fig 16; 154pp: English.
XX      This is PRO3434, a novel secreted protein. The invention
XX      concerns novel secreted and transmembrane proteins, designated PRO
XX      polypeptides. The cDNA and gene sequences are useful in the recombinant
XX      production of PRO polypeptides, as a hybridization probe to screen
XX      libraries to isolate cDNAs with sequence identity to PRO polypeptides or
XX      to map the gene encoding the PRO polypeptides and analyzing genetic
XX      disorders. The cDNA/gene can also be used to produce transgenic animals
XX      useful for the development and screening of therapeutically useful
XX      reagents. They can also be used in gene therapy, e.g. to replace a
XX      defective gene.
XX      Sequence 1029 AA:
XX      Query Match 100.0%; Score 1029; DB 21; Length 1029;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      181 QLRVLGPEDDLAMGFLQIFPLSPDRMOSSSPRPVALALQALGQELARVVOGSEFVPGI 240
Qy      181 QLRVLGPEDDLAMGFLQIFPLSPDRMOSSSPRPVALALQALGQELARVVOGSEFVPGI 240
Qy      241 TVRVLAALATLLSPHGALVMSMRSHFLACPLRLQCLQYRCYRPODTGESSLELKYLL 300
Db      241 TVRVLAALATLLSPHGALVMSMRSHFLACPLRLQCLQYRCYRPODTGESSLELKYLL 300
Qy      301 QMLQWLDSPGVGGGFLRAQRLMLASQASAGRLSDVRGGLRLAELAFRODLEVVSTV 360
Db      301 QMLQWLDSPGVGGGFLRAQRLMLASQASAGRLSDVRGGLRLAELAFRODLEVVSTV 360
Qy      361 RAVIATLRSGECCSVPEPDLISKVLOGLIEVRSPHLEELLTAFFSATADAASFPAACKPVY 420
Db      361 RAVIATLRSGECCSVPEPDLISKVLOGLIEVRSPHLEELLTAFFSATADAASFPAACKPVY 420
Qy      421 VVSSILLQEEEPFLAGCKRADGGSLEAVRLGSSGLVDVMDLEMDPEVYSSCPDLQLRL 480
Db      421 VVSSILLQEEEPFLAGCKRADGGSLEAVRLGSSGLVDVMDLEMDPEVYSSCPDLQLRL 480
Qy      481 FSRRKKGQAGVPSFRPYLTLFTHQSSWPTLHQCIRVLLGKSRQRPDPASLDPLMAC 540
Db      481 FSRRKKGQAGVPSFRPYLTLFTHQSSWPTLHQCIRVLLGKSRQRPDPASLDPLMAC 540
Qy      541 IHVPRIMQGRDQRTPOKRREELVLRVQGPBELISVELLAEETRSDGDTAAGSLIQAR 600
Db      541 IHVPRIMQGRDQRTPOKRREELVLRVQGPBELISVELLAEETRSDGDTAAGSLIQAR 600
Qy      601 LPLLLSCCGDDESVRKVTENHSGCIQMGDSVLRRCRDLLLOLYLQREELRVVPEVL 660
Db      601 LPLLLSCCGDDESVRKVTENHSGCIQMGDSVLRRCRDLLLOLYLQREELRVVPEVL 660
Qy      661 LHSEGAASSVCKLDGLIHREFTLLADTSDSRALENRGADASMACRKLAAVAPHLRLRL 720
Db      661 LHSEGAASSVCKLDGLIHREFTLLADTSDSRALENRGADASMACRKLAAVAPHLRLRL 720
Qy      721 PWIAALLHGRTHLNFQEFROQNHLSGFLHVLGLLELLOPHVFRSEHOGALWCCLISFIRL 780
Db      721 PWIAALLHGRTHLNFQEFROQNHLSGFLHVLGLLELLOPHVFRSEHOGALWCCLISFIRL 780
Qy      781 LILYKSSRHLLAFLNKFQFTHKITYTYPAPASIFLOKHADPLDLSFSDNSDIWLKSL 840
Db      781 LILYKSSRHLLAFLNKFQFTHKITYTYPAPASIFLOKHADPLDLSFSDNSDIWLKSL 840
Qy      841 LAGLSLPSRDRDTDGLDEEGEESAGSLPVYSLFTPLTAEMAPYMKRLSKQOTVE 900
Db      841 LAGLSLPSRDRDTDGLDEEGEESAGSLPVYSLFTPLTAEMAPYMKRLSKQOTVE 900
Qy      901 DLLEVLSDIDEMSRRPETLSFSTNLQRLMSABECCRNLAFLSALRSMQNSPSTIAAF 960
Db      901 DLLEVLSDIDEMSRRPETLSFSTNLQRLMSABECCRNLAFLSALRSMQNSPSTIAAF 960
Qy      961 LPTFWYCGSDQFEVYVQTRNLPEYALICQHAHVILHRAFLVGMYGMDPSAQISEAL 1020
Db      961 LPTFWYCGSDQFEVYVQTRNLPEYALICQHAHVILHRAFLVGMYGMDPSAQISEAL 1020
Qy      1021 RLHMEAVM 1029
Db      1021 RLHMEAVM 1029

```

RESULT 3  
 ID AU077765 standard; Protein: 1029 AA.  
 XX AU077765;  
 AC 05-JUN-2002 (first entry)  
 DT Tumour associated antigenic target polypeptide (TAT) 156.  
 DE Tumour associated antigenic target polypeptide: TAT; cancer:  
 KW breast cancer; colorectal cancer; lung cancer; ovarian cancer;  
 central nervous system cancer; liver cancer; bladder cancer;  
 KW

KW pancreatic cancer; cervical cancer; melanoma; leukemia; TAI156.  
XX Homo sapiens.  
XX Key  
FH Peptide  
FT Location/Qualifiers  
FT 1..16  
FT /label= Signal\_peptide  
FT 17..1029  
FT /label= Mature\_TAI156  
FT /note= "Tumour associated antigenic target  
FT polypeptide"  
FT Region  
FT 91..97  
FT /label= N-myristoylation\_site  
FT 136..142  
FT /label= N-myristoylation\_site  
FT 154..158  
FT /note= "CAMP and CGMP-dependent protein kinase  
FT phosphorylation site"  
FT Region  
FT 224..230  
FT /label= N-myristoylation\_site  
FT 329..333  
FT /label= Amidation\_site  
FT 331..335  
FT /note= "CAMP and CGMP-dependent protein kinase  
FT phosphorylation site"  
FT Region  
FT 435..441  
FT /label= N-myristoylation\_site  
FT 439..445  
FT /label= N-myristoylation\_site  
FT 443..449  
FT /label= N-myristoylation\_site  
FT 616..620  
FT /note= "CAMP and CGMP-dependent protein kinase  
FT phosphorylation site"  
FT Region  
FT 634..638  
FT /label= Amidation\_site  
FT 665..671  
FT /label= N-myristoylation\_site  
FT 698..704  
FT /label= N-myristoylation\_site  
FT 785..789  
FT /note= "CAMP and CGMP-dependent protein kinase  
FT phosphorylation site"  
FT Region  
FT 891..895  
FT /note= "CAMP and CGMP-dependent protein kinase  
FT phosphorylation site"  
FT WO200216581-A2.  
XX 28-FEB-2002.  
PD 14-AUG-2001: 2001WO-US25464.  
XX 24-AUG-2000: 2000WO-US23328.  
PR 28-FEB-2001: 2001WO-US06520.  
PR 22-JUN-2001: 2001US-0888257.  
PR 22-JUN-2001: 2001WO-US20118.  
XX (GETH ) GENENTECH INC.  
PA Gao W, Polakis P, Shou J, Smith V, Soriano R, Williams PM, Wu TD;  
PI Zhang Z;  
PI WPI: 2002-280928/32.  
DR N-PSDB; ABR11743.  
XX Novel isolated antibody which binds to tumour-associated antigenic  
PT target polypeptide useful for killing cancer cells expressing the  
PT polypeptide and for treating tumour comprising cells that expresses the  
PT polypeptide -  
XX Claim 2; Fig 7; 123pp; English.  
XX

CC The invention describes an isolated antibody which binds to a  
CC tumour-associated antigenic target (TAT) polypeptide. The antibody is  
CC useful for: killing a cancer cell (such as a breast, colorectal, lung,  
CC ovarian, central nervous system, liver, bladder, pancreatic, cervical,  
CC melanoma or leukemia cell) that expresses a polypeptide with at least  
CC 80% identity to the TAT polypeptide sequence; treating a tumour  
CC comprising cells that express a polypeptide with at least 80% identity  
CC to the TAT polypeptide sequence; determining the presence of a  
CC polypeptide having at least 80 % identity to the TAT polypeptide  
CC in a sample suspected of containing the polypeptide; diagnosing the  
CC presence of a tumour in a mammal, and for antibody dependent enzyme  
CC mediated prodrug therapy (ADEPT). This is the amino acid sequence of  
CC the tumour associated antigenic target polypeptide (TAT) 156, described  
CC in the invention.  
XX  
XX  
SQ Sequence 1029 AA;

Query Match 100.0%; Score 1029; DB 23; Length 1029;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHILVYHAWYILTLTGPPRADSEFOALDIWPEEKPLPTATLVPTSEBALPDMKL 60  
DB 1 MHILVYHAWYILTLTGPPRADSEFOALDIWPEEKPLPTATLVPTSEBALPDMKL 60  
QY 61 RMIRSEVLRLVDAALDLEPOQLLFVOSFGIPVSSMSKLLQFLDQAVAHDPQTLQONIM 120  
DB 61 RMIRSEVLRLVDAALDLEPOQLLFVOSFGIPVSSMSKLLQFLDQAVAHDPQTLQONIM 120  
QY 121 DKRYMAHVEVQHERGASGGQTFHSLTASLPRRSTAPKPKSSPEQITGGRIYVGT 180  
DB 121 DKRYMAHVEVQHERGASGGQTFHSLTASLPRRSTAPKPKSSPEQITGGRIYVGT 180  
QY 181 QLRVLGPEDDLQAGMFIQIFPLSPDPMOSSPPRVALLAQALQGLARVQSSPEVPGI 240  
DB 181 QLRVLGPEDDLQAGMFIQIFPLSPDPMOSSPPRVALLAQALQGLARVQSSPEVPGI 240  
QY 241 TVRVLOALATLSSPHGALVMSMRSHPLACPLRLQCYQRCVQDTGFSSLFKVL 300  
DB 241 TVRVLOALATLSSPHGALVMSMRSHPLACPLRLQCYQRCVQDTGFSSLFKVL 300  
QY 301 OMLOWLDSGVEGEPRLAOLRLMASQASAGRRLSDVRGGILRLAEALAFRODEEVSSVY 360  
DB 301 OMLOWLDSGVEGEPRLAOLRLMASQASAGRRLSDVRGGILRLAEALAFRODEEVSSVY 360  
QY 361 RAVIATLRSGEQSVPEDLISKVLQGLIEVRSPHEBELLTAFPSATADAASPPACKPVY 420  
DB 361 RAVIATLRSGEQSVPEDLISKVLQGLIEVRSPHEBELLTAFPSATADAASPPACKPVY 420  
QY 421 VVSSLLQEEBEPPLAGKPGADGSLFAVRLGESSGLLVMLDPEVYSSCPDLORL 480  
DB 421 VVSSLLQEEBEPPLAGKPGADGSLFAVRLGESSGLLVMLDPEVYSSCPDLORL 480  
QY 481 FSRKRGKGOAVSPFRPYLLTFTHOSWPTLHQICRYLLGKSRQDFPSASLDLTMK 540  
DB 481 FSRKRGKGOAVSPFRPYLLTFTHOSWPTLHQICRYLLGKSRQDFPSASLDLTMK 540  
QY 541 IHVPRTMOGRDQTPQKRREBELVRYOGPELISVELLAEATRSQDDTACSLIQAR 600  
DB 541 IHVPRTMOGRDQTPQKRREBELVRYOGPELISVELLAEATRSQDDTACSLIQAR 600  
QY 601 LPLLLSCCGGDESVKVTBHLISGCIQWGDVYLRGRCDLLQLYLRPELRVPYEV 660  
DB 601 LPLLLSCCGGDESVKVTBHLISGCIQWGDVYLRGRCDLLQLYLRPELRVPYEV 660  
QY 661 LHSEGAASSVCKLDGLIHREITLLADTSDSRALERMGADASAGACKLVAHPILLRL 720  
DB 661 LHSEGAASSVCKLDGLIHREITLLADTSDSRALERMGADASAGACKLVAHPILLRL 720  
QY 721 PMIAALLHGRTHLNFQEFQONHLSCLVHLGLLELLQHVRSERHOGALMDCLLSFIRL 780  
DB 721 PMIAALLHGRTHLNFQEFQONHLSCLVHLGLLELLQHVRSERHOGALMDCLLSFIRL 780

```
QY 781 LNNRKSRRHLAAFINKEVQFIHKYITYNAPAAISFLQKHADPLHLSFDNSDLVMLKSL 840
DB 781 LNNRKSRRHLAAFINKEVQFIHKYITYNAPAAISFLQKHADPLHLSFDNSDLVMLKSL 840
QY 841 LAGSLSPRRDRDTGDLDEEGEESASGLPLVSYSLFPLPLAAMAPYMKRLSGQIVE 900
DB 841 LAGSLSPRRDRDTGDLDEEGEESASGLPLVSYSLFPLPLAAMAPYMKRLSGQIVE 900
QY 901 DLLEVLSDIDEMSRRRPELISFSTNLRQMSAECCRNLAFLSLRSMONSPTAAAF 960
DB 901 DLLEVLSDIDEMSRRRPELISFSTNLRQMSAECCRNLAFLSLRSMONSPTAAAF 960
QY 961 LPTFWYCAGSODEFVYQALRNLPXYALLCQEHAAVLLHRAFLVGMGQMPDPAJISEAL 1020
DB 961 LPTFWYCAGSODEFVYQALRNLPXYALLCQEHAAVLLHRAFLVGMGQMPDPAJISEAL 1020
QY 1021 RLHMEAVM 1029
DB 1021 RLHMEAVM 1029

RESULT 4
ID AAB42495 standard; Protein: 650 AA.
AC AAB42495;
XX 08-FEB-2001 (first entry)
DE Human ORFX ORF2259 polypeptide sequence SEQ ID NO:4518.
XX
KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
KW vultenar; antiposrotic; antiparkinsonian; noctropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiatic;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineumatic; antihypertensive;
KW antiviral; antibacterial; antifungal; antineumatic; antihypertensive;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN MO200058473-A2.
XX
PD 05-OCT-2000.
XX
PE 31-MAR-2000; 2000MO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkels RA, Leach M;
XX
DR WPI: 2000-602362/57.
DR N-PSDB: AAC76704.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 3717-3718; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
```

```
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytosolic; hepatotropic; vultenar;
CC antiposrotic; antiparkinsonian; noctropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiatic; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antineumatic; antibacterial; antiviral; antifungal; antineumatic;
CC antihypertensive; antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 650 AA:
XX
Query Match 49.0%; Score 504; DB 21; Length 650;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 397 ELTAFPSATADASPFPACKPVVYSSLLDEEPLAGKRGAGGSIENVRIGPSSGL 456
DB 8 ELTAFPSATADASPFPACKPVVYSSLLDEEPLAGKRGAGGSIENVRIGPSSGL 67
QY 457 LVDWLEMLDPEVYSSCPDLQLRLFSRRKGGGOAOPSPRPULLTLFTHSSMPTLHOCI 516
DB 68 LVDWLEMLDPEVYSSCPDLQLRLFSRRKGGGOAOPSPRPULLTLFTHSSMPTLHOCI 127
QY 517 RVLLGKSRRQRPDPSASLDFIACIHVPRIWGRDQRPQKRREELVLRVGPPELISLVE 576
DB 128 RVLLGKSRRQRPDPSASLDFIACIHVPRIWGRDQRPQKRREELVLRVGPPELISLVE 187
QY 577 LTLAEFRSQGQDPAACSLIOARLPULLSCCGGDESVKRYTEHLSCGTOQWGSVYLR 636
DB 188 LTLAEFRSQGQDPAACSLIOARLPULLSCCGGDESVKRYTEHLSCGTOQWGSVYLR 247
QY 637 RCRDLLLOLYLQRPRLRVPVPEVLLHSEGAASSVCKLDGLIHRFTLLADTSDSRALEN 696
DB 248 RCRDLLLOLYLQRPRLRVPVPEVLLHSEGAASSVCKLDGLIHRFTLLADTSDSRALEN 307
QY 697 RGADASMACRKLAVAHPLLLRLHPLMIALHGRTHLNFQEPFRQNHLSGFLHVLGLLEL 756
DB 308 RGADASMACRKLAVAHPLLLRLHPLMIALHGRTHLNFQEPFRQNHLSGFLHVLGLLEL 367
QY 757 LQPHVFRSEHOGALMDCLISFIRLLNRRKSSRHAAAFINKVQFIHKYITYNAPAAISF 816
DB 368 LQPHVFRSEHOGALMDCLISFIRLLNRRKSSRHAAAFINKVQFIHKYITYNAPAAISF 427
QY 817 LQKHADPLHLSFDNSDLVMLKSLAGLSLSPRRDRDTGDLDEEGEESASGLPLVSY 876
DB 428 LQKHADPLHLSFDNSDLVMLKSLAGLSLSPRRDRDTGDLDEEGEESASGLPLVSY 487
QY 877 LFTPLTAEMADPYMKRLSGQIVE 900
DB 488 LFTPLTAEMADPYMKRLSGQIVE 511

RESULT 5
ID AAY87196 standard; Protein: 404 AA.
AC AAY87196;
XX 09-MAY-2000 (first entry)
XX
DE Human secreted protein sequence SEQ ID NO:235.
```



CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 71 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).  
 CC The present sequence represents a human secreted protein fragment  
 CC referred to in the disclosure of the invention.

CC Sequence 404 AA:

CC SQ

Query Match 39.3%; Score 404; DB 22; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 626 IIOQMGDSVIGRCRDLLOLYLQRELPVPVEVLHSGAASSVCKIDGLHRRITLL 685  
 DB 1 IIOQMGDSVIGRCRDLLOLYLQRELPVPVEVLHSGAASSVCKIDGLHRRITLL 60  
 OY 686 ADTSRRALENGADASMACRKLAVAHPLLLRHLPMTAALHGRTHLNFQFROONHLS 745  
 DB 61 ADTSRRALENGADASMACRKLAVAHPLLLRHLPMTAALHGRTHLNFQFROONHLS 120  
 OY 746 CETHLVGLLELLQPHVFNSEHOGALMDCILSTRILLNTRKSSRHAAINKEFVQFIHKY 805  
 DB 121 CETHLVGLLELLQPHVFNSEHOGALMDCILSTRILLNTRKSSRHAAINKEFVQFIHKY 180  
 OY 806 IYYNPAAPISFLOKHADPLHDSFNSDPLVMKSLAGLSLSPDRDRTRGDEEEERS 865  
 DB 181 IYYNPAAPISFLOKHADPLHDSFNSDPLVMKSLAGLSLSPDRDRTRGDEEEERS 240  
 OY 866 SAGSLPLVSVSLFTPLTAEMAPYMKRSLRGQTVEDLLEVLSDIDEMSRREPILISFEST 925  
 DB 241 SAGSLPLVSVSLFTPLTAEMAPYMKRSLRGQTVEDLLEVLSDIDEMSRREPILISFEST 300  
 OY 926 NIQRLMSSAECCRNLAFLARSMQNSISIAAEFLPTMYCLGSDFEVYOTALRNLP 985  
 DB 301 NIQRLMSSAECCRNLAFLARSMQNSISIAAEFLPTMYCLGSDFEVYOTALRNLP 360  
 OY 986 YALLCOEHAVALHRAFLVGMGTGMDPSQOISEALRIILMEAYM 1029  
 DB 361 YALLCOEHAVALHRAFLVGMGTGMDPSQOISEALRIILMEAYM 404

RESULT 7

ABG33997 ID ABG33997 standard; Protein; 404 AA.

XX ABG33997;

XX 15-JUL-2002 (first entry)

DE Human secreted protein #2 encoded by gene 46.

XX Human, secreted protein; gene therapy; immunosuppressive;

KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;

KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;  
 KW rhinoid; fungicide; ophthalmological; autoimmune disease; neoplasm;  
 KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;  
 KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;  
 KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;  
 KW ocular disorder; corneal infection; wound healing; skin aging;  
 KW epithelial cell proliferation; food additive.

OS Homo sapiens.

PN WO200226931-A2.

XX 04-APR-2002.

PF 24-SEP-2001; 2001WO-US29871.

XX 25-SEP-2000; 2000US-234925P.

PR 12-JAN-2001; 2001WO-US00911.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;  
 PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;  
 PI Endress GA, Mucenski M, Ebner R;

XX WPI: 2002-362489/39.

DR N-PSDB: ABK69771.

XX Novel 71 isolated secreted polypeptides and polynucleotides encoding

PT the polypeptides, useful for treating Huntington's disease, sepsis,  
 PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,  
 PT asthma

PS Disclosure: Page 1313-1314; 1478pp; English.

CC The invention relates to an isolated nucleic acid molecule (or its  
 CC fragment, homologue complement or allelic variant) encoding a human  
 CC secreted protein (and its fragment, domain, epitope, variant, secreted  
 CC form and species variant). Also included are a recombinant vector  
 CC comprising the nucleic acid, a recombinant host cell comprising the  
 CC vector, an antibody against the secreted protein, a recombinant host cell  
 CC that expresses the secreted protein and a method of identifying a binding  
 CC partner of the secreted protein. The nucleic acid and protein are used to  
 CC prevent, diagnose, treat or ameliorate a medical condition in e.g.  
 CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep  
 CC for example autoimmune diseases e.g. rheumatoid arthritis.  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver.  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders.  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
 CC ocular disorders e.g. corneal infection. Many other diseases and  
 CC disorders are listed in the specification. The polypeptides can also be  
 CC used to aid wound healing an epithelial cell proliferation, to prevent  
 CC skin aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis. The polypeptides can also be used as a food additive or  
 CC preservative to increase or decrease storage capabilities. The present  
 CC sequence represents a novel human secreted protein of the invention.

XX Sequence 404 AA:

Query Match 39.3%; Score 404; DB 23; Length 404;

Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 626 IIOQMGDSVIGRCRDLLOLYLQRELPVPVEVLHSGAASSVCKIDGLHRRITLL 685

DB 1 IIOQMGDSVIGRCRDLLOLYLQRELPVPVEVLHSGAASSVCKIDGLHRRITLL 60

OY 686 ADTSRRALENGADASMACRKLAVAHPLLLRHLPMTAALHGRTHLNFQFROONHLS 745

DB 61 ADTSRRALENGADASMACRKLAVAHPLLLRHLPMTAALHGRTHLNFQFROONHLS 120

QY 746 CFIHVLGELLELPHPYFRSRHOGALMDCLLSFIRLLINRKSSRHIAAFINKFVQFIHKY 805  
 DB 121 CFIHVLGELLELPHPYFRSRHOGALMDCLLSFIRLLINRKSSRHIAAFINKFVQFIHKY 180  
 QY 806 ITYNAPAAISFLQKHADPLHDLSDNSDLYMLKSLAGLSLPSRDRDTGDLDEGEES 865  
 DB 181 ITYNAPAAISFLQKHADPLHDLSDNSDLYMLKSLAGLSLPSRDRDTGDLDEGEES 240  
 QY 866 SAGSLPVSYSLEFPTLTAEMADPYMKRLSRGQTVEDLLEVLSDIDEMSRRPETLSFST 925  
 DB 241 SAGSLPVSYSLEFPTLTAEMADPYMKRLSRGQTVEDLLEVLSDIDEMSRRPETLSFST 300  
 QY 926 NIGRLSSAECCRNLAFLSIALRSMNSPSIAAFLPTFYKCLGSDQFEVVOALRLNLP 985  
 DB 301 NIGRLSSAECCRNLAFLSIALRSMNSPSIAAFLPTFYKCLGSDQFEVVOALRLNLP 360  
 QY 986 YALLCOEHAVALHRAFLVGMGQMPDPSAQISEALRIHMEAVM 1029  
 DB 361 YALLCOEHAVALHRAFLVGMGQMPDPSAQISEALRIHMEAVM 404

RESULT 8  
 AA06086  
 ID AA06086 standard; protein; 327 AA.  
 AC AA06086;  
 DT 24-SEP-2001 (first entry)  
 DE Human gene 46 encoded secreted protein HUBC35, SEQ ID NO:148.  
 XX  
 KW Human: secreted protein; proliferative disorder; cancer; tumour; asthma;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;  
 KW parotitis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;  
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;  
 KW angiotensin disorder; kidney disorder; gastrointestinal disorder; allergy;  
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;  
 KW cell culture; chemotaxis; vulnery; binding partner identification;  
 KW gene therapy.  
 XX  
 OS Homo sapiens.  
 EH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT /label= Signal\_peptide  
 FT 24..327  
 FT /label= Mature\_human\_secreted\_protein  
 PN WO200151504-A1.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 12-JAN-2001: 2001MO-US00911.  
 XX  
 PR 13-JAN-2000: 2000US-0482273.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;  
 PI Lafleur DM, Olsens HS, Brewer LA, Florence KA, Young PE, Soppet DR;  
 PI Endress GA, Muscenski M, Edner R;  
 XX  
 DR N-PDB: AAD11675.  
 XX  
 DR WPI: 2001-425865/45.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition  
 XX  
 PS Claim 11: Page 771-772: 864pp; English.  
 XX  
 CC AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted

CC protein genes, and AA06041-AA06132 represent the proteins they encode.  
 CC AA06133-AA06205 represent human secreted protein fragments.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 71 genes, and  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).  
 CC The present sequence represents a human secreted protein of  
 CC the invention.  
 CC  
 SO Sequence 327 AA:  
 XX  
 Query Match 31.8%; Score 327; DB 22; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 14e-314;  
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 703 MACKKLAVALHRAFLVGMGQMPDPSAQISEALRIHMEAVM 762  
 DB 1 MACKKLAVALHRAFLVGMGQMPDPSAQISEALRIHMEAVM 60  
 QY 763 RSEHOGALMDCLLSFIRLLINRKSSRHIAAFINKFVQFIHKYITYNAPAAISFLQKHAD 822  
 DB 61 RSEHOGALMDCLLSFIRLLINRKSSRHIAAFINKFVQFIHKYITYNAPAAISFLQKHAD 120  
 QY 823 PLHDLSDNSDLYMLKSLAGLSLPSRDRDTGDLDEGEESAGSLPVSYSLEFPTLT 882  
 DB 121 PLHDLSDNSDLYMLKSLAGLSLPSRDRDTGDLDEGEESAGSLPVSYSLEFPTLT 180  
 QY 883 AAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRRPETLSFSTNLQRLMSAECCRNLA 942  
 DB 181 AAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRRPETLSFSTNLQRLMSAECCRNLA 240  
 QY 943 FSLALRSMNSPSIAAFLPTFYKCLGSDQFEVVOALRLNPEYALLCOEHAVALHRAFL 1002  
 DB 241 FSLALRSMNSPSIAAFLPTFYKCLGSDQFEVVOALRLNPEYALLCOEHAVALHRAFL 300  
 QY 1003 LVGMGQMPDPSAQISEALRIHMEAVM 1029  
 DB 301 LVGMGQMPDPSAQISEALRIHMEAVM 327

RESULT 9  
 ABG33908  
 ID ABG33908 standard; protein; 327 AA.  
 AC ABG33908;  
 DT 15-JUL-2002 (first entry)  
 DE Human secreted protein encode by gene 46.  
 XX  
 DE Human secreted protein encode by gene 46.  
 XX  
 KW Human: secreted protein; gene therapy; immunosuppressive;  
 KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;  
 KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;

KM virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;  
 KM rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;  
 KM cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;  
 KM angiodysplasia; nervous system disorder; Alzheimer's disease; infection;  
 KM ocular disorder; corneal infection; wound healing; skin aging;  
 KM epithelial cell proliferation; food additive.  
 OS Homo sapiens.  
 XX WO200226931-A2.  
 XX 04-APR-2002.  
 XX 24-SEP-2001; 2001WO-US29871.  
 XX 25-SEP-2000; 2000US-234925P.  
 XX 12-JAN-2001; 2001WO-US00911.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;  
 PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;  
 PI Endress GA, Mucenski M, Edner R;  
 XX WPI; 2002-362489/39.  
 XX N-PSDB; AAK69771.  
 XX Novel 71 isolated secreted polypeptides and polynucleotides encoding  
 PT the polypeptides, useful for treating Huntington's disease, sepsis,  
 PT meningitis, thrombocytopenia, haemolytic anaemia, rheumatoid arthritis,  
 PT asthma  
 XX Claim 11: Page 1250-1251; 1478pp; English.  
 XX The invention relates to an isolated nucleic acid molecule (or its  
 CC fragment, homologue complement or allelic variant) encoding a human  
 CC secreted protein (and its fragment, domain, epitope, variant, secreted  
 CC form and species variant). Also included are a recombinant vector  
 CC comprising the nucleic acid, a recombinant host cell comprising the  
 CC vector, an antibody against the secreted protein, a recombinant host cell  
 CC that expresses the secreted protein and a method of identifying a binding  
 CC partner of the secreted protein. The nucleic acid and protein are used to  
 CC prevent, diagnose, treat or ameliorate a medical condition in e.g.  
 CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep  
 CC for example autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiodysplasia, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
 CC ocular disorders e.g. corneal infection. Many other diseases and  
 CC disorders are listed in the specification. The polypeptides can also be  
 CC used to aid wound healing an epithelial cell proliferation, to prevent  
 CC skin aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis. The polypeptides can also be used as a food additive or  
 CC preservative to increase or decrease storage capabilities. The present  
 CC sequence represents a novel human secreted protein of the invention.  
 XX  
 SQ Sequence 327 AA:  
 Query Match 31.8%; Score 327; DB 23; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-314;  
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 121 PLHDSFSDMSDLVMLKSLAGLSLPSRDDRTDGLDEGEESAGSLPLVSVSLPTPLT 180  
 QY 883 AAEAPYMRRLSRGQVEVDLEVLSDIDEMSRRRPILSFPSNLORLMSSAECCRNLA 942  
 DB 181 AAEAPYMRRLSRGQVEVDLEVLSDIDEMSRRRPILSFPSNLORLMSSAECCRNLA 240  
 QY 943 FSIALRSMONSPSIAAAPTFTFCGSDQFEVYQALNLPYALLCOEHAVALHRAE 1002  
 DB 241 FSIALRSMONSPSIAAAPTFTFTFCGSDQFEVYQALNLPYALLCOEHAVALHRAE 300  
 QY 1003 LVGMYGQMDPSAQISEALRLILHMEAVM 1029  
 DB 301 LVGMYGQMDPSAQISEALRLILHMEAVM 327  
 RESULT 10  
 AAY87109  
 ID AAY87109 standard; Protein; 328 AA.  
 XX  
 AC AAY87109;  
 DT 09-MAY-2000 (first entry)  
 XX  
 DE Human secreted protein sequence SEQ ID NO:148.  
 XX  
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
 KW antiinflammatory; nootropic; neuroprotective; antiallergic; cancer;  
 KW tumour; neurodegenerative disorder; developmental abnormality; allergy;  
 KW foetal deficiency; blood disorder; immune system disorder; arthritis;  
 KW autoimmune disease; hepatic disease; renal disease; inflammation;  
 KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;  
 KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;  
 KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;  
 KW reproductive disorder; gastrointestinal disorder; respiratory disorder;  
 KW metabolic disorder; food additive; preservative.  
 XX  
 XX Homo sapiens.  
 XX WO200004140-A1.  
 XX 27-JAN-2000.  
 XX 14-JUL-1999; 99WO-US15849.  
 XX 15-JUL-1998; 98US-0092921.  
 XX 15-JUL-1998; 98US-0092922.  
 XX 15-JUL-1998; 98US-0092956.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;  
 PI Lafleur DW, Edner R, Olsen HS, Brewer LA, Florence KA, Young PE;  
 PI Mucenski M, Endress GA, Soppet DR;  
 XX WPI; 2000-161128/14.  
 XX N-PSDB; AAZ98062.  
 XX New isolated human genes, useful for diagnosis and treatment of, e.g.  
 PT cancers, neurological or blood disorders  
 XX  
 XX Claim 11: Page 404-406; 494pp; English.  
 XX The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the  
 CC human secreted proteins given in AAY87064 to AAY87223. Human secreted  
 CC protein can have activities based on the tissues and cells the genes are  
 CC expressed in. Examples of activities include: cytostatic;  
 CC immunosuppressive; antiinflammatory; nootropic; neuroprotective; and  
 CC antiallergic. The polynucleotides and their corresponding secreted  
 CC polypeptides are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g. by protein or gene therapy. Also pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the new



CC polynucleotides. Human secreted protein s and their polynucleotides can  
CC be used for developing products for the diagnosis or treatment of cancer  
CC tumours, neurodegenerative disorders, developmental abnormalities and  
CC foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, hepatic and renal disease, inflammation,  
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,  
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,  
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,  
CC cardiovascular disorders, reproductive disorders, gastrointestinal  
CC disorders, respiratory disorders and metabolic disorders. The  
CC proteins or polynucleotides can also be used as food additives or  
CC preservatives. The proteins are also useful for identifying their  
CC binding partners. AAZ8008 to AAZ9016 and AAY87063 are sequence used in  
CC the exemplification of the present invention.

**SQ Sequence 328 AA;**

Query Match	31.8%	Score 327	DB 21	length 328
Best Local Similarity	100.0%	Pred. No.	1.4e-314	
Matches 327	Conservative 0	Mismatches 0	Indels 0	Gaps 0

[illegible]

**OY**    763 RSEHOGALWDCLLSFIRLLLNTRKSSRHIAAFINKEVQFIHKYITYNAPAAISFLQKHAD 822  
**Db**     61 RSEHOGALWDCLLSFIRLLLNTRKSSRHIAAFINKEVQFIHKYITYNAPAAISFLQKHAD 122

**QY** 823 PLHDSFDNSDLVLMKSLLAGLSLPSRDRTDRGLDEEGEESSAGSLPLVSYSLLFTPLT 88  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
**Db** 121 PLHDSFDNSDLVLMKSLLAGLSLPSRDRTDRGLDEEGEESSAGSLPLVSYSLLFTPLT 180

Qy 883 AAEMAPYMKRSLRGQTVEDDLLEVLSDIDMSRRREILSFSTNTQLRMSSAECCRNLA 94  
|||  
Db 181 AAEMAPYMKRSLRGQTVEDDLLEVLSDIDMSRRREILSFSTNTQLRMSSAECCRNLA 24

```

09      FSLLRSMONSPSIAAFLPTFWYCLGSDQFEVQTALRNUPETALLCQEHAAVLLHRAF 100
      |||||||
241  FSLLRSMONSPSIAAFLPTFWYCLGSDQFEVQTALRNUPETALLCQEHAAVLLHRAF 300

```

Qy	1003	LVGMYGOMPSAQISEALRILHMEAVM	1029
Db	301	LVGMYGOMPSAQISEALRILHMEAVM	327

RESULT 11	.
AAM95497	
ID	AAM95497 standard; Protein; 161 AA
XX	

AC	AAM95497;
XX	
DT	21-NOV-2001 (first entry)
XX	

Accession	Protein	Gene	Organism	Accession	Protein	Gene	Organism
DE	Human reproductive system related antigen SEQ ID NO: 4155.			DE	Human reproductive system related antigen SEQ ID NO: 4155.		
XX				XX			
KW	Human; reproductive system related antigen; reproductive s			KW	Human; reproductive system related antigen; reproductive s		
KW	Cancer; gene therapy			KW	Cancer; gene therapy		

OS	Homo sapiens.
XX	
PN	W0200155320-A2.
XY	

PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01339.
XY	

PR 11-JAN-2000; 2000US-0179065  
PR 04-FEB-2000; 2000US-0180628  
PR 24-FEB-2000; 2000US-0184664  
PR 02-MAR-2000; 2000US-0186350  
PR 16-MAR-2000; 2000US-0189874  
PR 17-MAR-2000; 2000US-0190076

PR	18-APR-2000	2000US-0198123	PR	18-APR-2000	2000US-0198123
PR	19-MAY-2000	2000US-0205515	PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209467	PR	07-JUN-2000	2000US-0209467
PR	26-JUN-2000	2000US-0214886	PR	26-JUN-2000	2000US-0214886
PR	30-JUN-2000	2000US-0215135	PR	30-JUN-2000	2000US-0215135
PR	07-JUL-2000	2000US-0216647	PR	07-JUL-2000	2000US-0216647
PR	07-JUL-2000	2000US-0216680	PR	07-JUL-2000	2000US-0216680
PR	11-JUL-2000	2000US-0217487	PR	11-JUL-2000	2000US-0217487
PR	14-JUL-2000	2000US-0218290	PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220964	PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518	PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224570	PR	14-AUG-2000	2000US-0224570
PR	14-AUG-2000	2000US-0225447	PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225757	PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225762	PR	14-AUG-2000	2000US-0225762
PR	14-AUG-2000	2000US-0225767	PR	14-AUG-2000	2000US-0225767
PR	14-AUG-2000	2000US-0225568	PR	14-AUG-2000	2000US-0225568
PR	14-AUG-2000	2000US-0225470	PR	14-AUG-2000	2000US-0225470
PR	14-AUG-2000	2000US-0225477	PR	14-AUG-2000	2000US-0225477
PR	22-AUG-2000	2000US-0226681	PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226688	PR	22-AUG-2000	2000US-0226688
PR	22-AUG-2000	2000US-0227182	PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227824	PR	23-AUG-2000	2000US-0227824
PR	30-SEP-2000	2000US-0228920	PR	30-SEP-2000	2000US-0228920
PR	01-SEP-2000	2000US-0229287	PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343	PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344	PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345	PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509	PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513	PR	05-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0229517	PR	05-SEP-2000	2000US-0229517
PR	06-SEP-2000	2000US-0230436	PR	06-SEP-2000	2000US-0230436
PR	06-SEP-2000	2000US-0230437	PR	06-SEP-2000	2000US-0230437
PR	08-SEP-2000	2000US-0231142	PR	08-SEP-2000	2000US-0231142
PR	08-SEP-2000	2000US-0231143	PR	08-SEP-2000	2000US-0231143
PR	08-SEP-2000	2000US-0231144	PR	08-SEP-2000	2000US-0231144
PR	08-SEP-2000	2000US-0231145	PR	08-SEP-2000	2000US-0231145
PR	08-SEP-2000	2000US-0231146	PR	08-SEP-2000	2000US-0231146
PR	08-SEP-2000	2000US-0231147	PR	08-SEP-2000	2000US-0231147
PR	08-SEP-2000	2000US-0232081	PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0233168	PR	12-SEP-2000	2000US-0233168
PR	14-SEP-2000	2000US-0233297	PR	14-SEP-2000	2000US-0233297
PR	14-SEP-2000	2000US-0233365	PR	14-SEP-2000	2000US-0233365
PR	14-SEP-2000	2000US-0233398	PR	14-SEP-2000	2000US-0233398
PR	21-SEP-2000	2000US-0234423	PR	21-SEP-2000	2000US-0234423
PR	21-SEP-2000	2000US-0234424	PR	21-SEP-2000	2000US-0234424
PR	25-SEP-2000	2000US-0234497	PR	25-SEP-2000	2000US-0234497
PR	25-SEP-2000	2000US-0234498	PR	25-SEP-2000	2000US-0234498
PR	26-SEP-2000	2000US-0234884	PR	26-SEP-2000	2000US-0234884
PR	27-SEP-2000	2000US-0235834	PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836	PR	27-SEP-2000	2000US-0235836
PR	29				



PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249266.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 06-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 RA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR MPI; 2001-465570/50.  
 DR N-PSDB; AAL01467.  
 XX  
 XX Isolated nucleic acid molecule encoding a reproductive system antigen  
 PT is used in preventing, treating or ameliorating a medical condition  
 XX  
 PS Claim 11; SEQ ID NO 4155; 1297pp + Sequence Listing; English.  
 CC  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human reproductive system related antigens. These can be used  
 CC in the prevention and treatment of reproductive system disorders,  
 CC

CC including cancer. The present sequence is a protein of the invention.  
 XX  
 SO Sequence 161 AA;  
 CC  
 CC Query Match 4.2%; Score 43; DB 22; Length 161;  
 CC Best Local Similarity 100.0%; Pred. No. 9,4e-34;  
 CC Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC 851 DRTDGLDEEGEESSAGSLPLVSVLFTPLTAAMAPYMKRL 893  
 CC  
 CC Db 41 DRTDGLDEEGEESSAGSLPLVSVLFTPLTAAMAPYMKRL 83  
 CC  
 CC RESULT 12  
 CC ID AAU52278 standard; Protein; 55 AA.  
 CC  
 CC AC AAU52278;  
 CC  
 CC DT 27-FEB-2002 (first entry)  
 CC  
 CC DE Propionibacterium acnes immunogenic protein #13174.  
 CC  
 CC KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 CC KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 CC KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 CC KW dermatological; osteopathic; neuroprotectant.  
 CC  
 CC OS Propionibacterium acnes.  
 CC  
 CC PN WO200181581-A2.  
 CC  
 CC PD 01-NOV-2001.  
 CC  
 CC PF 20-APR-2001; 2001WO-US12865.  
 CC  
 CC PR 21-APR-2000; 2000US-199047P.  
 CC PR 02-JUN-2000; 2000US-208841P.  
 CC PR 07-JUL-2000; 2000US-216747P.  
 CC  
 CC XX  
 CC PA (CORI-) CORIXA CORP.  
 CC  
 CC XX Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 CC PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 CC XX MPI; 2001-616774/71.  
 CC DR N-PSDB; AAS59553.  
 CC  
 CC XX  
 CC PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 CC PT vaccinating against and diagnosing infections, especially useful for  
 CC PT treating acne vulgaris  
 CC  
 CC PS Example 1; SEQ ID No 13473; 1069pp; English.  
 CC  
 CC XX  
 CC CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC CC P. acnes is also involved in infections of bone, joints and the central  
 CC CC nervous system, however it is particularly involved in the inflammatory  
 CC CC lesions associated with acne vulgaris. A method for detecting the  
 CC CC presence or absence of P. acnes in a patient comprises contacting a  
 CC CC sample with a binding agent that binds to the proteins of the invention  
 CC CC and determining the amount of bound protein in the sample. The  
 CC CC polypeptides may be used as antigens in the production of antibodies  
 CC CC specific for P. acnes proteins. These antibodies can be used to  
 CC CC downregulate expression and activity of P. acnes polypeptides and  
 CC CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC CC diagnostic agents for determining P. acnes presence, for example, by  
 CC CC enzyme linked immunosorbent assay (ELISA).  
 CC CC Note: The sequence data for this patent did not form part of the printed  
 CC CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 55 AA;  
 Query Match 0.8%; Score 8; DB 22; Length 55;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 331 RRLSDVNG 338  
 |||||  
 Db 12 RRLSDVNG 19

RESULT 13.  
 AAU64883  
 ID AAU64883 standard; Protein; 84 AA.  
 XX  
 AC AAU64883;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #25779.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN MO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001MO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 DR N-PSDB; AAS59654.  
 DR WPI; 2001-616774/71.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID NO 26078; 10699P; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 84 AA;  
 Query Match 0.8%; Score 8; DB 22; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 405 ATADASP 412  
 |||||  
 Db 3 ATADASP 10

RESULT 14  
 ABP34284  
 ID ABP34284 standard; Protein; 101 AA.  
 XX  
 AC ABP34284;  
 XX  
 DT 09-JUL-2002 (first entry)  
 XX  
 DE Human kinase-like ORF3257 protein, SEQ ID NO:6514.  
 XX  
 KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; actinin; inhibitor; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vunerary;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antihypoid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200190366-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 24-MAY-2001; 2001MO-US17076.  
 XX  
 PR 24-MAY-2000; 2000US-206690P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Leach MD, Shinkels RA;  
 PI WPI; 2002-106200/14.  
 DR N-PSDB; ABN78310.  
 DR Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation -  
 XX  
 PS Claim 10; Page 1875-1876; 2508PP; English.  
 XX  
 CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,

CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokine activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antifibrotic activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
 CC  
 XX  
 SQ Sequence 101 AA:

Query Match 0.8%; Score 8; DB 23; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 750 VGLLELL 757  
 |||||  
 DB 36 VGLLELL 43

## RESULT 15

ABB58449  
 ID ABB58449 standard; Protein; 207 AA.

XX  
 AC ABB58449;

XX  
 DT 26-MAR-2002 (first entry)

XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 2139.

XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 pharmaceutical.

XX  
 OS Drosophila melanogaster.

XX  
 PN WO200171042-A2.

XX  
 PD 27-SEP-2001.

XX  
 PF 23-MAR-2001; 2001WO-US09231.

XX  
 PR 23-MAR-2000; 2000US-191637P.

XX  
 PR 11-JUL-2000; 2000US-0614150.

XX  
 PA (PEKE ) PE CORP NY.

XX  
 PI Venter JC, Adams M, LI PWD, Myers EW;

XX  
 DR WPI; 2001-656860/75.

XX  
 DR N-PSDB; ABL02552.

XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

XX  
 PS Disclosure; SEQ ID NO 2139; 21pp + Sequence Listing; English.

XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX

SQ Sequence 207 AA:

Query Match 0.8%; Score 8; DB 22; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 427 LOEEPELA 434  
 |||||  
 DB 145 LOEEPELA 152

Search completed: May 7, 2003, 19:44:31  
 Job time : 47 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:43:23 ; Search time 27 Seconds  
(without alignments)  
3663.790 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 1029  
Sequence: 1 MHILVHAMVILTLTGPPRA.....MDPSAQISEALRIILMEAVM 1029

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues  
Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  
1: PIR-73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	633	61.5	640	2 T08758	hypothetical prote
2	100	9.7	310	2 T46429	hypothetical prote
3	9	0.9	188	2 G75326	probable 16S rRNA
4	9	0.9	241	2 B75628	DNA-binding stress
5	9	0.9	327	2 A83436	hypothetical prote
6	8	0.8	157	2 T24380	hypothetical prote
7	8	0.8	272	2 E90194	hypothetical prote
8	8	0.8	279	2 G95117	Hemk protein [limpo
9	8	0.8	345	2 E88103	protein W10G11.5 [
10	8	0.8	386	2 G89801	hypothetical prote
11	8	0.8	410	2 A29393	hemocyanin beta-c
12	8	0.8	423	2 A70947	hypothetical prote
13	8	0.8	593	2 AB0046	probable N-acetyl
14	8	0.8	640	2 F75553	hypothetical prote
15	8	0.8	958	2 C87504	hypothetical prote
16	8	0.8	997	2 T32814	hypothetical prote
17	8	0.8	1154	2 A13431	hypothetical prote
18	8	0.8	1815	2 S73021	chromosome segrega
19	8	0.8	1822	2 F87203	polyketide synthas
20	8	0.7	53	2 B82583	hypothetical prote
21	8	0.7	88	2 A61356	hypothetical prote
22	7	0.7	88	2 A11726	hypothetical prote
23	7	0.7	89	2 A83465	flagellar biosynth
24	7	0.7	103	2 H71134	hypothetical prote
25	7	0.7	111	2 AC3411	hypothetical cytos
26	7	0.7	113	2 AB0736	probable bacteriop
27	7	0.7	128	2 E70866	probable globin -
28	7	0.7	132	2 P00060	T-cell receptor be
29	7	0.7	148	2 T37069	hypothetical prote

30	7	0.7	150	2 D69100	probable protein d
31	7	0.7	156	2 B59225	allergen Bos d 2.0
32	7	0.7	156	2 A59225	allergen Bos d 2.0
33	7	0.7	156	2 A75149	hypothetical prote
34	7	0.7	160	2 B84903	hypothetical prote
35	7	0.7	163	2 S14269	phenomene-binding
36	7	0.7	171	2 E84912	probable polyribon
37	7	0.7	183	2 F71127	hypothetical prote
38	7	0.7	185	2 A72528	hypothetical prote
39	7	0.7	188	2 T02247	hypothetical prote
40	7	0.7	194	2 T46920	hypothetical prote
41	7	0.7	196	2 E69100	hypothetical prote
42	7	0.7	197	2 F65019	hypothetical prote
43	7	0.7	198	2 B83267	probable glutathio
44	7	0.7	200	2 JN0579	DNA-binding protei
45	7	0.7	204	2 AF0914	conserved hypothet

ALIGNMENTS

RESULT 1										
T08758										
hypothetical protein DKFZP586J0619.1 - human (fragment)										
C:Species: Homo sapiens (man)										
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999										
C:Accession: T08758										
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.										
submitted to the Protein Sequence Database, March 1999										
A:Reference number: Z16471										
A:Accession: T08758										
A:Molecule type: mRNA										
A:Residues: 1-640 <MAN>										
A:Cross-references: EMBL:AL050110										
A:Experimental source: adult uterus; clone DKFZP586J0619										
A:Genetics:										
A:Note: DKFZP586J0619.1										
Query Match										
Best Local Similarity 61.5%; Score 633; DB 2; Length 640;										
Best Local Similarity 100.0%; Pred. No. 0;										
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;										
QY	397	ELTTAFSSATADAA	SPACKPVVVVSSLLQEEEP	LAGKPGADGSL	EAVALGSSGL	456				
DB	8	ELTTAFSSATADAA	SPACKPVVVVSSLLQEEEP	LAGKPGADGSL	EAVALGSSGL	67				
QY	457	LVDWLEMLDPEVYSSCPDLQLRL	FSRRKRGKGAQVPSFRPYLLTLFT	HOSSMPTLHOCI	516					
DB	68	LVDWLEMLDPEVYSSCPDLQLRL	FSRRKRGKGAQVPSFRPYLLTLFT	HOSSMPTLHOCI	127					
QY	517	RVLLGKSREQRFPDSASDLFL	MACIHVPRIWQGRDQRT	POKRREELVLRVOGPELISLVE	576					
DB	128	RVLLGKSREQRFPDSASDLFL	MACIHVPRIWQGRDQRT	POKRREELVLRVOGPELISLVE	187					
QY	577	LILAEATRSQDGDGTAAACSLIQARL	PLLSCCGDDES	SVKVTENHLSGCIQMGDSVLGR	636					
DB	188	LILAEATRSQDGDGTAAACSLIQARL	PLLSCCGDDES	SVKVTENHLSGCIQMGDSVLGR	247					
QY	637	KCRDLLQLYLQREBELRV	PPVEVLLHSEGA	SSVCKDLGLIHRFTLLADPS	SDALEN	696				
DB	248	KCRDLLQLYLQREBELRV	PPVEVLLHSEGA	SSVCKDLGLIHRFTLLADPS	DALEN	307				
QY	697	RGADASMCRKLA	VAHPDLLRLPLMIAL	LGRTILNFQEEFRQNHLS	CFHLVGLLEL	756				
DB	308	RGADASMCRKLA	VAHPDLLRLPLMIAL	LGRTILNFQEEFRQNHLS	CFHLVGLLEL	367				
QY	757	LQPHVFRSEHOGALWDCILS	FTRLIN	YRKSSRH	LAFTNKEVQFTLHKYIT	NAIPAISF	816			
DB	368	LQPHVFRSEHOGALWDCILS	FTRLIN	YRKSSRH	LAFTNKEVQFTLHKYIT	NAIPAISF	427			
QY	817	LQKHADPLHDI	SFDNSDI	YMLKSL	LAGLSL	PSRDDRTD	GDGDEE	GEESASG	SLPLVSVS	876
DB	428	LQKHADPLHDI	SFDNSDI	YMLKSL	LAGLSL	PSRDDRTD	GDGDEE	GEESASG	SLPLVSVS	487

QY 877 LFTPLTAEMAPYMKRLSRGQIVEDLLEVLSDIDEMSRREPELISFSTNLQRLMSAAE 936  
 DB 488 LFTPLTAEMAPYMKRLSRGQIVEDLLEVLSDIDEMSRREPELISFSTNLQRLMSAAE 547  
 QY 937 CCNLFALFSLARFMONSPSTAAAFLEPFMYCLISODEEVQVTLARNPEVALLCOEHAAV 996  
 DB 548 CCNLFALFSLARFMONSPSTAAAFLEPFMYCLISODEEVQVTLARNPEVALLCOEHAAV 607  
 QY 997 LHRRAFLVGMVGMQDPSAQISEALRLIMEAVM 1029  
 DB 608 LHRRAFLVGMVGMQDPSAQISEALRLIMEAVM 640  
 RESULT 2  
 T6429  
 hypothetical protein DKFZp434C0126.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T6429  
 R:Ansgorge, W.; Winkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z33028  
 A:Accession: T6429  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-310 <AAA>  
 A:Cross-references: EMBL:AL137358  
 A:Experimental source: adult testis; clone DKFZp434C0126  
 C:Genetics:  
 A:Note: DKFZp434C0126.1  
 Query Match 9.7%; Score 100; DB 2; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-92;  
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 844 LSLPSRDDRTDRCIDEGEESSAGSLPLVSVSLFPTLTAEMAPYMKRLSRGQIVEDLL 903  
 DB 89 LSLPSRDDRTDRCIDEGEESSAGSLPLVSVSLFPTLTAEMAPYMKRLSRGQIVEDLL 148  
 QY 904 EVLSDIDEMSRREPELISFSTNLQRLMSAAECCRNLA 943  
 DB 149 EVLSDIDEMSRREPELISFSTNLQRLMSAAECCRNLA 188  
 RESULT 3  
 G75326  
 probable 16S rRNA processing protein RlmM - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: G75326  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MID:20036896; PMID:10567266  
 A:Accession: G75326  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-188 <WHI>  
 A:Cross-references: GB:AE002038; GB:AE000513; MID:96459790; PIDN:AAF1157.1; PID:9645979  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR2010  
 A:Map position: 1  
 Query Match 0.9%; Score 9; DB 2; Length 188;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 432 PLAGKPGA 440  
 |||||||

DB 155 PLAGKPGA 163  
 RESULT 4  
 B75628  
 DNA-binding stress response protein, Dps family - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: B75628  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MID:20036896; PMID:10567266  
 A:Accession: B75628  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-241 <WHI>  
 A:Cross-references: GB:AE001826; MID:96460827; PIDN:AAF12541.1; PID:96460837; TIGR:DR  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRB0092  
 A:Map position: megaplasmid  
 A:Genome: plasmid  
 A:Note: plasmid MPI  
 Query Match 0.9%; Score 9; DB 2; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 419 VVVVSSILL 427  
 DB 8 VVVVSSILL 16  
 RESULT 5  
 AB3436  
 hypothetical protein PA1680 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: AB3436  
 R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
 Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; L  
 D.; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: AB2950; MID:20437337; PMID:10984043  
 A:Accession: AB3436  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-327 <STO>  
 A:Cross-references: GB:AE004595; GB:AE004091; MID:99947644; PIDN:AA05069.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA1680  
 Query Match 0.9%; Score 9; DB 2; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 340 LRLRAEALA 348  
 DB 81 LRLRAEALA 89  
 RESULT 6  
 T24380  
 hypothetical protein T03D8.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T24380  
 R:Montmore, B.

submitted to the EMBL Data Library, March 1997  
 A:Reference number: 219884  
 A:Accession: T24380  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-157 <WIL>  
 A:Cross-references: EMBL:292838; PIDN:CAB07406.1; GSPDB:GN00023; CESP:T03D8.2  
 A:Experimental source: clone T03D8  
 C:Genetics:  
 A:Gene: CESP:T03D8.2  
 A:Map position: 5  
 A:Introns: 49/3; 83/1; 125/3

Query Match 0.8%; Score 8; DB 2; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 TLETHQSS 508  
 DB 27 TLETHQSS 34  
 |||||

RESULT 7  
 E90194  
 hypothetical protein psta [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 A:Accession: E90194  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.  
 arett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: E90194  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-272 <KUR>  
 A:Cross-references: GB:AE006641; NID:g13813648; PIDN:AAK40812.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: psta

Query Match 0.8%; Score 8; DB 2; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1015 QISEALRI 1022  
 DB 152 QISEALRI 159  
 |||||

RESULT 8  
 G95117  
 HemK protein [imported] - Streptococcus pneumoniae (strain TIGR4)  
 C:Species: Streptococcus pneumoniae  
 C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
 A:Accession: G95117  
 R:Petelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid-  
 on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Kadane, D.; Holtzap-  
 pson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
 A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.  
 A:Reference number: A95000; MUID:21357209; PMID:11463916  
 A:Accession: G95117  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-279 <KUR>  
 A:Cross-references: GB:AE005672; PIDN:AAK75136.1; PID:g14972494; GSPDB:GN00164; TIGR:SP4  
 C:Experimental source: strain TIGR4  
 C:Genetics:  
 A:Gene: SP1021

Query Match 0.8%; Score 8; DB 2; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 LVELLIAE 581  
 DB 99 LVELLIAE 106  
 |||||

RESULT 9  
 E88103  
 protein W10G11.5 [imported] - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
 A:Accession: E88103  
 R:Anonymous, The C. elegans Sequencing Consortium.  
 Science 282, 2012-2018, 1998  
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio-  
 A:Reference number: A75000; MUID:99069613; PMID:9851916  
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-  
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;  
 A:Accession: E88103  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-345 <STO>  
 A:Cross-references: GB:chr\_II; PIDN:AB95070.1; PID:g2746928; GSPDB:GN00020; CESP:W10G  
 A:Gene: W10G11.5  
 A:Map position: 2

Query Match 0.8%; Score 8; DB 2; Length 345;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 TLLADTSD 690  
 DB 50 TLLADTSD 57  
 |||||

RESULT 10  
 G89801  
 hypothetical protein SA0346 [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 A:Accession: G89801  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O-  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: AB9758; MUID:21311952; PMID:11418146  
 A:Accession: G89801  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-386 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13700272; PIDN:BAR41570.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA0346  
 C:Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 0.8%; Score 8; DB 2; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 LPLVSVSL 877  
 DB 312 LPLVSVSL 319  
 |||||

RESULT 11  
 A29393  
 hemocyanin beta-c chain d - Roman snail

C:Species: Helix pomatia (Roman snail)  
C:Date: 13-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 11-Jan-2000  
C:Accession: A29393  
R:Prexell, R.; Slegmund, S.; Schneider, H.J.; Linzen, B.; Gielens, C.; Preaux, G.; Lontje  
Biol. Chem. Hoppe-Seyler 368, 617-635, 1987  
A:Title: Complete amino-acid sequence of a functional unit from a molluscan hemocyanin  
A:Reference number: A29393; MVID:87299002; PMID:3620107  
A:Accession: A29393  
A:Molecule type: protein  
A:Residues: 1-410 <DRE>  
C:Superfamily: hemocyanin

Query Match 0.8%; Score 8; DB 2; Length 410;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 523 SREORFDP 530  
|||||  
DB 115 SREORFDP 122

## RESULT 12

A70947  
hypothetical protein RV3163c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: A70947

R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
Connor, R.; Davies, R.; Fellin, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MVID:98295987; PMID:9634230

A:Accession: A70947

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-423 <COL>

A:Cross-references: GB:AL021646; GB:AL123456; NID:93242278; PIDN:CAA16628.1; PID:9282757

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3163c

Query Match 0.8%; Score 8; DB 2; Length 423;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 975 VVOTALRN 982  
|||||  
DB 258 VVOTALRN 265

## RESULT 13

AB0046  
probable N-acetylmuramoyl-L-alanine amidase-family protein YP00370 [imported] - Yersinia

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 17-May-2002

C:Accession: AB0046

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchall, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
H. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, S.

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MVID:21470413; PMID:11586360

A:Accession: AB0046

A:Molecule type: preliminary

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-593 <KUR>

A:Cross-references: GB:AL590842; PIDN:CA089229.1; PID:915978468; GSPDB:GN00175

C:Genetics:

A:Gene: YP00370

C:Superfamily: Escherichia coli hypothetical protein b2817

Query Match 0.8%; Score 8; DB 2; Length 593;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 417 KPVVWSS 424  
|||||  
DB 168 KPVVWSS 175

## RESULT 14

F75553  
hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 26-May-2000

C:Accession: F75553

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M.; Shen, M.; Yamatehyan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MVID:20036896; PMID:10567266

A:Accession: F75553

A:Molecule type: preliminary

A:Status: preliminary

A:Residues: 1-640 <WHI>

A:Cross-references: GB:AE001878; GB:AE000513; NID:96457820; PIDN:AAF09751.1; PID:9645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0160

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DR0160

Query Match 0.8%; Score 8; DB 2; Length 640;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 840 LLAGLSLP 847  
|||||  
DB 372 LLAGLSLP 379

## RESULT 15

C87504  
hypothetical protein CC2058 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: C87504

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.;  
Laub, M.T.; Debay, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MVID:21173698; PMID:11259647

A:Accession: C87504

A:Molecule type: preliminary

A:Status: preliminary

A:Residues: 1-958 <STO>

A:Cross-references: GB:AE005673; NID:913423537; PIDN:AAK24031.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC2058

Query Match 0.8%; Score 8; DB 2; Length 958;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 328 SAGRRISD 335  
|||||  
DB 665 SAGRRISD 672

Search completed: May 7, 2003, 19:46:23  
Job time: 33 secs



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:42:23 ; Search time 15 seconds  
(without alignments)  
2845.276 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 1029  
Sequence: 1 MHILVHAWYILLTLGPPRA.....MDPSAQSSEALRIHMEAVM 1029

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	9	188	1 R1MM_DEIRA	O9rsu1 delinococcus
2	8	335	1 LEP4_MYXA	O30387 myxococcus
3	8	338	1 CYF_PHOA	P93522 phormidium
4	8	410	1 HCYB_HELO	P12031 helix pomati
5	7	128	1 GIBO_MYCTU	O53197 mycobacteri
6	7	163	1 BPPI_AMPPE	O17077 anthraxeae p
7	7	172	1 ALI2_BOVIN	O28133 bos taurus
8	7	197	1 YRFP_ECOLI	P75547 escherichia
9	7	207	1 RECR_RALSO	O8y050 ralstonia s
10	7	208	1 SPC3_STRPU	P16537 strongyloce
11	7	209	1 ENMD_ECOS7	O8xbw8 escherichia
12	7	209	1 ENMD_ECOS7	P19925 escherichia
13	7	209	1 ENMD_ECOS7	O54153 shigella fl
14	7	211	1 THIE_BACHD	O9kcy8 bacillus ha
15	7	233	1 Y33L_SYNT3	P72583 synchocyst
16	7	244	1 HIS4_BACHD	O9k625 bacillus ha
17	7	245	1 OCCM_RHME	P72296 rhizobium m
18	7	258	1 NORC_NEIMA	O91vco neisseria m
19	7	262	1 APP6_ALLAR	P50363 allomyces a
20	7	262	1 APP6_ALLAR	P50364 allomyces m
21	7	264	1 COMT_RAT	P22734 rattus norv
22	7	265	1 COMT_MOUSE	O88587 mus musculu
23	7	270	1 RPA2_MOUSE	O62193 mus musculu
24	7	285	1 YX47_STRCO	O9x8h0 streptomyce
25	7	287	1 DCMH_HYDPS	P19914 hydrogenoph
26	7	289	1 FRAH_ANASP	P46017 anabaena sp
27	7	289	1 KPPR_ECOLI	P37307 escherichia
28	7	290	1 VINT_BPL2	P42540 bacterioph
29	7	290	1 BCHL_RHOB	O918j5 rhodospiril
30	7	293	1 VNP7_CAEBL	P34560 caenorhabdi
31	7	294	1 SC17_PICCA	O9p4d0 pichia past
32	7	299	1 YRFE_ECOLI	P75826 escherichia
33	7	300	1 HAP2_KLUDI	P53768 kluyveromyc

34	7	0.7	302	1	CCDC_CAEBL	P17657 caenorhabdi
35	7	0.7	302	1	GBPR_AGRU	P25547 agrobacteri
36	7	0.7	303	1	NUIM_CABUN	O78700 cabassou u
37	7	0.7	304	1	HEVI_CANFA	O9rsz2 canis famli
38	7	0.7	308	1	PYRB_PYROO	O58451 pyrococcus
39	7	0.7	317	1	STAL_HUMAN	P21815 homo sapien
40	7	0.7	320	1	YCF5_ECOLI	P75954 escherichia
41	7	0.7	326	1	YNS6_CAEBL	P34590 caenorhabdi
42	7	0.7	329	1	MENC_HAEIN	P44961 haemophilus
43	7	0.7	331	1	DEFA_HUMAN	O00273 homo sapien
44	7	0.7	333	1	CXCI_HUMAN	P46094 homo sapien
45	7	0.7	347	1	F16P_YEAST	P09201 saccharomyc

## ALIGNMENTS

RESULT 1  
R1MM\_DEIRA STANDARD; PRT; 188 AA.  
AC O9rsu1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable 16S rRNA processing protein r1mm.  
GN R1MM OR DR2010.  
OS Delinococcus radiodurans.  
OC Bacteria; Thermus/Delinococcus group; Delinococci; Delinococcales;  
OC Delinococcaceae; Delinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Halt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utechtack T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RA "Genome sequence of the radioresistant bacterium Delinococcus  
RT radiodurans RI.";  
RT Science 286:1571-1577(1999).  
RL -!- FUNCTION: ESSENTIAL FOR EFFICIENT PROCESSING OF 16S RNA. PROBABLY  
PART OF THE 30S SUBUNIT PRIOR TO OR DURING THE FINAL STEP IN THE  
PROCESSING OF 16S PRE-30S RIBOSOMAL SUBUNTS. IT COULD BE SOME  
ACCESSORY PROTEIN NEEDED FOR EFFICIENT ASSEMBLY OF THE 30S  
SUBUNIT. R1MM IS NEEDED IN A STEP PRIOR TO RBEA DURING THE  
MATURATION OF 16S RNA. HAS AFFINITY FOR FREE RIBOSOMAL 30S  
SUBUNTS BUT NOT FOR 70S RIBOSOMES (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (potential).  
CC -!- SIMILARITY: BELONGS TO THE R1MM FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isp.slb.ch/announce/>  
CC or send an email to [license@slb.slb.ch](mailto:license@slb.slb.ch)).  
CC -----  
CC EMBL: AE002038; AAF11557.1; -  
CC TIGR: DR2010; -  
DR InterPro: IPR002676; R1MM.  
DR Pfam: PF01782; R1MM; 1.  
KW rRNA processing; Complete proteome.  
SQ SEQUENCE 188 AA; 19686 MW; 49P582D781CC99FD CRC64;  
Query Match 0.9%; Score 9; DB 1; Length 188;  
Best Local Similarity 100.0%; Pred. No. 0.65;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 432 PLAGRPGA 440  
 DB 155 PLAGRPGA 163

## RESULT 2

LEP4\_MYXA

ID LEP4\_MYXA STANDARD: PRT: 335 AA.

AC 030387;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Type 4 prepillin-like proteins leader peptide processing enzyme

DE (includes: Leader peptidase (EC 3.4.99.-) (Pepillin peptidase); N-

DE methyltransferase (EC 2.1.1.-)).

OS PILD.

GN Myxococcus xanthus.

OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;

OC Myxococcales; Cytoobacteriales; Myxococcaceae; Myxococcus.

OX NCBI\_TaxID=34;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-DK1622;

RA Wu S.S., Cheng Y.L., Kaiser D.;

RT "The pilH gene encodes an ABC transporter required for type IV pilus

RT biogenesis in Myxococcus xanthus."

RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: CLEAVES TYPE-4 FIMBRIAL LEADER SEQUENCE AND METHYLATES

CC THE N-TERMINAL (GENERALLY PHE) RESIDUE (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A24.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: AF003632; AAC36157.1; -

DR MEROPS: A24.001; -

DR InterPro: IPR000045; Peptidase.C20.

DR Pfam: PF01478; Peptidase.C20.1.

DR PRINTS: PR00864; PREPLINPTASE.

DR MULTIFUNCTIONAL enzyme: Hydrolase; Protease; Transferase;

KW Methyltransferase; Transmembrane.

FT TRANSMEM 13 33 POTENTIAL.

FT TRANSMEM 105 125 POTENTIAL.

FT TRANSMEM 131 151 POTENTIAL.

FT TRANSMEM 206 226 POTENTIAL.

FT TRANSMEM 258 278 POTENTIAL.

FT TRANSMEM 299 319 POTENTIAL.

SQ SEQUENCE 335 AA; 36767 MW; 27DF58A5FD2819FA CRC64;

Query Match 0.8%; Score 8; DB 1; Length 335;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 857 LDEGEER 864

DB 283 LDEGEER 290

## RESULT 3

CYF\_PHOLA

ID CYF\_PHOLA STANDARD: PRT: 338 AA.

AC P95522;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Apocytochrome F precursor.

OS PETA.

GN

OS Phormidium lamosum.  
 OC Bacteria; Cyanobacteria; Oscillatoriales; Phormidium.  
 OX NCBI\_TaxID=32059;

## RESULT 4

HCYB\_HELPO

ID HCYB\_HELPO STANDARD: PRT: 410 AA.

AC P12031;

DT 01-OCT-1989 (Rel. 12, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Hemocyanin, beta-C chain unit D (Fragment).

DE Helix pomatia (Roman snail) (Edible snail).

OS Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylomatophora;

GN

RT cytochrome f from the

RT Biochemistry 38:9590-9599(1999).

CC -1- FUNCTION: TRANSLOCATES PROTONS ACROSS THE THYLAKOID MEMBRANE AND

CC TRANSFERS ELECTRONS FROM THE RIESKE IRON-SULFUR PROTEIN AND PASSES

CC THEM TO PLASTOCYANIN. THIS FUNCTION IS VERY SIMILAR TO THAT OF

CC MITOCHONDRIAL CYTOCHROME C1.

CC -1- SUBUNIT: MEMBER OF THE CYTOCHROME B6/F COMPLEX INCLUDING

CC CYTOCHROME B6, CYTOCHROME F AND PROBABLY AN IRON SULFUR PROTEIN.

CC -1- SUBCELLULAR LOCATION: THYLAKOID MEMBRANE (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: Y09612; CAAT0824.1; -

DR PDB: 1CI3; 31-MAY-00.

DR InterPro: IPR002325; Apocyt\_F.

DR Pfam: PF01333; Apocytochrome.F.1.

DR PRINTS: PS00180; CYTOCHROME.C.1.

DR PROSITE: PS00180; CYTOCHROME.C.1.

DR ELECTRON transport; Heme; Photosynthesis; Photosystem I;

KW Photosystem II; Signal; 3D-structure.

FT CHAIN 1 45

FT METAL 46 338 APOCYTOCHROME F.

FT BINDING 46 46 IRON (HEME AXIAL LIGAND).

FT BINDING 66 66 HEME (COVALENT).

FT BINDING 69 69 HEME (COVALENT).

FT METAL 70 70 IRON (HEME AXIAL LIGAND).

SQ SEQUENCE 338 AA; 36069 MW; BD3D52036181FB50 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 338;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 MWLLTLG 16

DB 23 MWLLTLG 30

OC Helicaceae; Helicidae; Helix.  
 OX NCBI\_TaxID=5336;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=87299002; PubMed=3620107;  
 RA Drexel R., Siegmund S., Schneider H.J., Linzen B., Gjelens C.,  
 RA Preaux G., Lontle R., Kellermann J., Lottspeich F.;  
 RT "Complete amino-acid sequence of a functional unit from a molluscan  
 RT haemocyanin (Helix pomatia).";  
 RL Biol. Chem. Hoppe-Seyler 368:617-633(1987).  
 RN [2]  
 RP PARTIAL SEQUENCE, THIOETHER BOND, REVISIONS, AND MASS SPECTROMETRY.  
 RX MEDLINE=98000289; PubMed=9342242;  
 RA Gjelens C., de Geest N., Xin X.-Q., Devreese B., van Beeumen J.,  
 RA Preaux G.;  
 RT "Evidence for a cysteine-histidine thioether bridge in functional  
 RT units of molluscan haemocyanins and location of the disulfide bridges  
 RT in functional units d and g of the beta-c-haemocyanin of Helix  
 RT pomatia";  
 RL Eur. J. Biochem. 248:879-888(1997).  
 CC -1- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS  
 CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND  
 CC ARTHROPODS.  
 CC -1- COFACTOR: CONTAINS TWO COPPER ATOMS PER FUNCTIONAL UNIT.  
 CC -1- SUBUNIT: DECAMERS OF LARGE IDENTICAL SUBUNITS (450 kDa), EACH  
 CC CONTAINING 8 GLOBULAR OXYGEN-BINDING FUNCTIONAL UNITS.  
 CC -1- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN  
 CC SUBFAMILY.  
 CC PTR: A29393; A29393.  
 DR InterPro: IPR002227; Tyrosinase.  
 DR Pfam: PF00264; Tyrosinase. 1.  
 DR PRINTS: PR00092; TYROSINASE.  
 DR PROSITE: PS00497; TYROSINASE\_1; 1.  
 DR PROSITE: PS00498; TYROSINASE\_2; 1.  
 KW Oxygen transport; Transport; Copper; Glycoprotein;  
 KW Hemolymph; Repeat.  
 FT METAL 1 1  
 FT METAL 44 44  
 FT METAL 55 55  
 FT METAL 71 71  
 FT METAL 175 175  
 FT METAL 179 179  
 FT METAL 206 206  
 FT DISULFID 50 59  
 FT DISULFID 165 232  
 FT DISULFID 321 332  
 FT THIOETH 60 62  
 FT CARBOHYD 253 253  
 FT CONFLICT 60 60  
 FT CONFLICT 62 62  
 FT CONFLICT 165 165  
 FT NON\_TER 410 410  
 SQ SEQUENCE 410 AA; 47019 MW; E872743ED1AE66BF CRC64;  
 Query Match 0.8%; Score 8; DB 1; Length 410;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 523 SREORFDP 530  
 DB 115 SREORFDP 122  
 ID GIBO\_MYCTU STANDARD; PRT; 128 AA.  
 AC 053197;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hemoglobin-like protein HBO.  
 GN GIBO OR RV2470 OR MT2546 OR MTV008.26.  
 OS Mycobacterium tuberculosis, and

OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773, 1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Baocock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham T., Gentes S., Hamlin N., Holtroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellon M.S., Squares S., Squares R.,  
 RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Hatt D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.bovis; STRAIN=BCG;  
 RA Pathania R., Ramande P., Dikshit K.L.;  
 RT "Cloning, expression and characterization of an oxygen-binding protein  
 RT encoded by gibo locus of Mycobacterium tuberculosis and Mycobacterium  
 RT bovis";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL021246; CAI16047.1; -  
 DR EMBL: AF007091; AAK46846.1; ALT\_INIT.  
 DR EMBL: AF213450; AAF36597.1; -  
 DR TIGR: MT2546; -  
 DR TuberculList; RV2470; -  
 DR InterPro: IPR001486; Bac\_globin.  
 DR ProDom: PD004840; Bac\_globin. 1.  
 DR PROSITE: PS01213; GLOBIN\_FAM.2; 1.  
 KW Heme; Oxygen transport; Transport; Complete proteome.  
 FT METAL 75 75  
 FT METAL 75 75  
 SQ SEQUENCE 128 AA; 14950 MW; CEF4CD23907B0CCA CRC64;  
 Query Match 0.7%; Score 7; DB 1; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 187 PEDDLAG 193  
 DB 37 PEDDLAG 43  
 ID PBPI\_ANTPE STANDARD; PRT; 163 AA.  
 AC 053197;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hemoglobin-like protein HBO.  
 GN GIBO OR RV2470 OR MT2546 OR MTV008.26.  
 OS Mycobacterium tuberculosis, and

AC 01707;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Pheromone-binding protein 1 precursor (PBP 1) (APR-1).  
 OS Antheraea pernyi (Chinese oak silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
 OC Ditrysia; Bombycoidea; Saturniidae; Saturniinae; Saturniini;  
 OC Antheraea.  
 NCBI\_TaxID=7119;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Antenna;  
 RX MEDLINE=91268328; PubMed=2097299;  
 RA Raming K., Krieger J., Breer H.;  
 RT Primary structure of a pheromone-binding protein from *Antheraea pernyi*: homologues with other ligand-carrying proteins.";  
 RL J. Comp. Physiol. B 160:503-509(1990).  
 CC -1- FUNCTION: THIS MAJOR SOLUBLE PROTEIN IN OLFACTORY SENSILLA OF MALE MOTHS MIGHT SERVE TO SOLUBILIZE THE EXTREMELY HYDROPHOBIC PHEROMONE MOLECULES AND TO TRANSPORT PHEROMONE THROUGH THE AQUEOUS LYMPH TO RECEPTORS LOCATED ON OLFACTORY CILIA.  
 CC -1- TISSUE SPECIFICITY: ANTENNA.  
 CC -1- SIMILARITY: BELONGS TO THE PBP/GOBP FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X96773; CAA65376.1; -  
 DR HSSP: P34174; IDOE.  
 DR InterPro: IPR000746; PBP\_GOBP\_family.  
 DR Pfam: PF01395; PBP\_GOBP.1.  
 DR PRINTS: PR00484; PBPGOBP.  
 KW Pheromone-binding; Pheromone response; Transport; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 163 PHEROMONE-BINDING PROTEIN 1.  
 SO SEQUENCE 163 AA; 18086 MW; 3CFE4FE3E369C56 CRC64;  
 QY 698 GADASMA 704  
 DB 102 GADASMA 108  
 Query Match 0.7%; Score 7; DB 1; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RT "Complementary DNA cloning of the predominant allergen of bovine dander: a new member in the lipocalin family.";  
 RT J. Allergy Clin. Immunol. 97:1297-1303(1996).  
 RN [2]  
 RP 3D-STRUCTURE MODELING.  
 RX MEDLINE=99068569; PubMed=9853680;  
 RA Santa H., Saarela J.T., Laatikainen R., Rautialainen J., Virtanen T., Rytkonen M., Maentylaeervi R.;  
 RT "A bovine dander allergen, comparative modeling, and similarities and differences in folding with related proteins.";  
 RT J. Protein Chem. 17:657-662(1998).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=99107890; PubMed=9891000;  
 RA Rouvainen J., Rautialainen J., Virtanen T., Zeller T., Kauppinen J., Taivainen A., Maentylaeervi R.;  
 RT "Probing the molecular basis of allergy, three-dimensional structure of the bovine lipocalin allergen bos d 2.";  
 RL J. Biol. Chem. 274:2337-2343(1999).  
 RN [4]  
 RP CHARACTERIZATION, AND CRYSTALLIZATION.  
 RX MEDLINE=98321199; PubMed=9647765;  
 RA Rautialainen J., Aurioja S., Rouvainen J., Kauppinen J., Zeller T., Virtanen T., Maentylaeervi R.;  
 RT "Tissue localization of bovine dander allergen Bos d 2.";  
 RL J. Allergy Clin. Immunol. 101:349-353(1998).  
 CC -1- FUNCTION: PROBABLE PHEROMONE CARRIER.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN SKIN. PRODUCED IN SWEAT GLANDS AND TRANSPORTED TO THE SKIN SURFACE.  
 CC -1- MISCELLANEOUS: PREDOMINANT ALLERGEN OF BOVINE DANDER.  
 CC -1- SIMILARITY: BELONGS TO THE LIPOCALIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: I42867; AAB08720.1; -  
 DR PDB: 1BU7; 1I-MAY-99.  
 DR InterPro: IPR000566; Lipocalin\_cytfABP.  
 DR Pfam: PF00061; Lipocalin.1.  
 DR PROSITE: PS00213; LIPOCALIN; FALSE-NEG.  
 KW Transport; Lipocalin; Signal; Allergen; 3D-structure.  
 FT SIGNAL 1 16  
 FT CHAIN 17 172 ALLERGEN BOS D 2.  
 FT MOD\_RES 17 17 PYRROLIDONE CARBOXYLIC ACID.  
 FT DISULFID 60 64  
 FT DISULFID 79 170  
 SO SEQUENCE 172 AA; 19560 MW; C50F66F1B3D98156 CRC64;  
 QY 311 VEGGPLR 317  
 DB 46 VEGGPLR 52  
 Query Match 0.7%; Score 7; DB 1; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8  
 YFPF\_ECOLI

```

ID YFPF_ECOLI STANDARD: PRT: 197 AA.
AC P76547;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfp.
GN YFPF OR B2447.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia
CC NCBI_TaxID=562;
RX NCBI_TaxID=562;
RN [1]
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000331; AAC75500.1; -
DR EcoGene; EG14177; yfp.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 197 AA; 22907 MW; 6E1B6EA72D49B519 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 821 ADPLHDL 827
DB 15 ADPLHDL 21

RESULT 9
RECR_RALSO STANDARD: PRT: 207 AA.
AC Q8Y050;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Recombination protein recr.
GN RECR OR RSC1194 OR RS05726.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
CC Ralstonia.
CC NCBI_TaxID=305;
RN [1]
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billaut A., Brotier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Gunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigler P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Welzenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
CC -1- FUNCTION: May play a role in DNA repair. It seems to be involved
CC in an reBC-independent recombinational process of DNA repair. It
CC may act with recr and reco (By similarity).
CC -1- SIMILARITY: BELONGS TO THE RECR FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL646063; CAD14896.1; -
DR InterPro; IPR002936; DNAPrim_toprim.
DR InterPro; IPR000093; Recr.
DR Pfam; PF01751; Toprim; 1.
DR Pfam; PF02132; Recr; 1.
DR SMART; SM00493; TOPRIM; 1.
DR TIGRfam; TIGR00615; recr; 1.
DR PROSITE; PS01300; RECR; FALSE_NEG.
KW DNA repair; DNA recombination; Zinc-finger; Complete proteome.
FT ZN_FING 62
FT C4_TYPE (POTENTIAL).
SQ SEQUENCE 207 AA; 22510 MW; 1428A8F817D03071 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 342 RLAEALA 348
DB 47 RLAEALA 53

RESULT 10
SPEC3_STRPU STANDARD: PRT: 208 AA.
AC P16537;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SPEC3 protein.
GN SPEC3.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidae; Strongylocentrotidae;
OC Strongylocentrotus.
CC NCBI_TaxID=7668;
RN [1]
RC STRAIN=7668;
RX MEDLINE=88112804; PubMed=2828169;
RA Eldon E.D., Angerer L.M., Angerer R.C., Klein W.H.;
RT "Spec3: embryonic expression of a sea urchin gene whose product is
RT involved in ectodermal ciliogenesis."
RL Genes Dev. 1:1280-1292(1987).
RN [2]
RC SUBCELLULAR LOCATION.
RX MEDLINE=90169510; PubMed=2407617;
RA Eldon E.D., Montpetit I.C., Nguyen T., Decker G., Valdivia M.C.,
RA Klein W.H., Brandhorst B.P.;
RT "Localization of the sea urchin Spec3 protein to cilia and Golgi
RT complexes of embryonic ectoderm cells."
RL Genes Dev. 4:111-122(1990).
CC -1- SUBCELLULAR LOCATION: IN CILIA AND GOLGI COMPLEXES.
DR PIR; A43696; A43696.
KW Transmembrane.
FT TRANSMEM 107
FT TRANSMEM 163
FT CAROXYD 71
FT N-LINKED (GLCNAc...) (POTENTIAL).
SQ SEQUENCE 208 AA; 21719 MW; 4BE0638F1DE65C2E CRC64;

Query Match 0.7%; Score 7; DB 1; Length 208;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 438 PCADGGS 444
DB 11111111

```

Db 138 PGADGS 144

RESULT 11

ENTD\_ECO57 STANDARD: PRT: 209 AA.

AC Q8XB8; 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 4'-phosphopantetheinyl transferase end (EC 2.7.8.-) (Enterobactin synthetase component D) (Enterobactin synthase D).

GN ENTD OR 20723 OR ECS0622.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7"; Nature 409:529-533(2001).

RL [2]

RN SEQUENCE FROM N.A.

RC STRAIN-O157:H7 / K1MD 0509952;

RX PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hatori M., Shingawa H.;

RA "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12."; DNA Res. 8:11-22(2001).

RL -1- FUNCTION: Catalyzes the transfer of the 4'-phosphopantetheine moiety from coenzyme A to apo-domains of both entB (an ATP domain) and entF (a PCP domain). Plays an essential role in the assembly of the enterobactin (By similarity).

CC -1- CATALYTIC ACTIVITY: CoA + apo-entB/F -> adenosine 3',5'-bisphosphate + holo-entB/F.

CC -1- COFACTOR: Magnesium (By similarity).

CC -1- PATHWAY: Enterobactin biosynthesis.

CC -1- SUBUNIT: EntB, entD, entE, and entF form a multienzyme complex called enterobactin synthetase (By similarity).

CC -1- SIMILARITY: BELONGS TO THE P-PANT TRANSFERASE SUPERFAMILY.

CC ENTD FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AE005238; AAG54917.1; -

DR EMBL: AF002352; BAB34045.1; -

DR InterPro: IPR002582; ACPS.

DR InterPro: IPR003542; Enbnc-synthetD.

DR Pfam: PF01648; ACPS. 1.

DR PRINTS: PR01399; ENTSTNTHASD.

KW Transferase; Enterobactin biosynthesis; Transport; Iron transport; Magnesium; Complete proteome.

FT METAL 110 110 MAGNESIUM (BY SIMILARITY).

FT METAL 112 112 MAGNESIUM (BY SIMILARITY).

FT METAL 155 155 MAGNESIUM (BY SIMILARITY).

SEQUENCE 209 AA; 23477 MW; 236960055E5765AE CRC64;

Query Match 0.7%; Score 7; DB 1; Length 209;

Best local Similarity 100.0%; Pred. No. 74;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 941 LAFSLAL 947

Db 141 LAFSLAL 147

RESULT 12

ENTD\_ECOLI STANDARD: PRT: 209 AA.

AC P19925; P77092;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 4'-phosphopantetheinyl transferase end (EC 2.7.8.-) (Enterobactin synthetase component D) (Enterobactin synthase D).

GN ENTD OR B0583.

OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE=90132534; PubMed=2533240;

RA Codere P.E., Earhart C.F.;

RT "The entD gene of the Escherichia coli K12 enterobactin gene cluster."; J. Gen. Microbiol. 135:3043-3055(1989).

RL [2]

RN ERRATUM.

RA Codere P.E., Earhart C.F.;

RL J. Gen. Microbiol. 136:1667-1667(1990).

RL [3]

RN SEQUENCE FROM N.A.

RP MEDLINE=89313305; PubMed=2526281;

RX Armstrong S.K., Pettis G.S., Forrester L.J., McIntosh M.A.;

RT "The Escherichia coli enterobactin biosynthesis gene, entD: nucleotide sequence and membrane localization of its protein product."; J. Gen. Microbiol. 3:757-766(1989).

RL [4]

RN SEQUENCE FROM N.A.

RP STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RA "The complete genome sequence of Escherichia coli K-12."; Science 277:1453-1474(1997).

RL [5]

RN SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE=97061202; PubMed=8905223;

RA Oshino T., Alha H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizochichi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiuchi T.;

RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map."; DNA Res. 3:137-155(1996).

RL [6]

RN SEQUENCE FROM N.A.

RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M., Federspiel N., Hyman R., Kaiman S., Komp C., Kudi O., Lew H., Lin D., Nannath A., Oefner P., Roberts D., Schramm S., Davis R.W.;

RA Submitted (JAN-1997) to the EMBL/GenBank/DBD databases.

```

RN [7]
RP FUNCTION.
RX PubMed-89929709;
RA Lambiot R.H., Gehring A.M., Flugel R.S., Zuber P., Lacelle M.,
RA Marahel M.A., Reid R., Khosla C., Walsh C.T.;
RT "A new enzyme superfamily - the phosphopantetheinyl transferases."
RL Chem. Biol. 3:923-936(1996).
RN [8]
RP FUNCTION.
RX STRAIN-BL21-DE3;
RX PubMed-9214294;
RA Gehring A.M., Bradley K.A., Walsh C.T.;
RT "Enterobactin biosynthesis in Escherichia coli: Isochorismate lyase
RT (Entb) is a bifunctional enzyme that is phosphopantetheinylated by
RT Entb and then acetylated by Entc using ATP and 2,3-dihydroxybenzoate."
RL Biochemistry 36:8495-8503(1997).
CC -1- FUNCTION: Catalyzes the transfer of the 4'-phosphopantetheine
CC moiety from coenzyme A to apo-domains of both entb (an ArCP
CC domain) and entf (a PCP domain). Plays an essential role in the
CC assembly of the enterobactin.
CC -1- CATALYTIC ACTIVITY: CoA + apo-entb/F = adenosine 3',5'-
CC bisphosphate + holo-entb/F.
CC -1- COFACTOR: Magnesium.
CC -1- PATHWAY: Enterobactin biosynthesis.
CC -1- SUBUNIT: Entb, entd, ente, and entf form a multienzyme complex
CC called enterobactin synthetase.
CC -1- SIMILARITY: BELONGS TO THE P-PANT TRANSFERASE SUPERFAMILY.
CC ENTD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17426; CAB57861.1; -
DR EMBL; AE000163; AAC73684.1; -
DR EMBL; D90700; BAA35224.1; ALT_INIT.
DR EMBL; U82598; AAB40782.1; ALT_INIT.
DR PIR; S06981; S06981.
DR EcGene; EG10262; entd.
DR InterPro; IPR002582; ACPS.
DR InterPro; IPR003542; Embac_synthetd.
DR Pfam; PF01648; ACPS. 1.
DR PRINTS; PR01399; ENTSNTHASED.
KW Transferase; Enterobactin biosynthesis; Transport; Iron transport;
KW Magnesium; Complete proteome.
FT METAL 110 110 MAGNESIUM (BY SIMILARITY).
FT METAL 112 112 MAGNESIUM (BY SIMILARITY).
FT METAL 155 155 MAGNESIUM (BY SIMILARITY).
FT METAL 155 155 MAGNESIUM (BY SIMILARITY).
FT CONFLICT 77 78 EL -> DV (IN REF. 3)
SQ SEQUENCE 209 AA; 23604 MW; B98403563D9DA7D CRC64;

Query Match 0.7%; Score 7; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 941 LAFSLAL 947
DB 141 LAFSLAL 147

```

```

GN ENTD.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RA Johansen K.A.;
RT "entd enterobactin biosynthesis genes of enteric bacteria."
RT Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
RL -1- FUNCTION: Catalyzes the transfer of the 4'-phosphopantetheine
CC moiety from coenzyme A to apo-domains of both entb (an ArCP
CC domain) and entf (a PCP domain). Plays an essential role in the
CC assembly of the enterobactin (By similarity).
CC -1- CATALYTIC ACTIVITY: CoA + apo-entb/F = adenosine 3',5'-
CC bisphosphate + holo-entb/F.
CC -1- COFACTOR: Magnesium (By similarity).
CC -1- PATHWAY: Enterobactin biosynthesis.
CC -1- SUBUNIT: Entb, entd, ente, and entf form a multienzyme
CC complex called enterobactin synthetase (By similarity).
CC -1- SIMILARITY: BELONGS TO THE P-PANT TRANSFERASE SUPERFAMILY.
CC ENTD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U52684; AAA97936.1; -
DR InterPro; IPR002582; ACPS.
DR InterPro; IPR003542; Embac_synthetd.
DR Pfam; PF01648; ACPS. 1.
DR PRINTS; PR01399; ENTSNTHASED.
KW Transferase; Enterobactin biosynthesis; Transport; Iron transport;
KW Magnesium.
FT METAL 110 110 MAGNESIUM (BY SIMILARITY).
FT METAL 112 112 MAGNESIUM (BY SIMILARITY).
FT METAL 155 155 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 209 AA; 23487 MW; 72F2B8C05C19C3E CRC64;

Query Match 0.7%; Score 7; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 941 LAFSLAL 947
DB 141 LAFSLAL 147

```

```

RESULT 13
ENTD_SHIFL STANDARD: PRT; 209 AA.
AC Q54153;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 4'-phosphopantetheinyl transferase entd (EC 2.7.8.-) (Enterobactin
DE synthetase component D) (Enterobactin synthase D).

```

```

RESULT 14
THIE_BACHD STANDARD: PRT; 211 AA.
AC Q9KCY8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thiamine-phosphate pyrophosphorylase (EC 2.5.1.3) (TMP
DE pyrophosphorylase) (TMP-Pase) (Thiamine-phosphate synthase).
GN GN THIE OR BH1431.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kihara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

```

```

RT  halodurans and genomic sequence comparison with Bacillus subtilis.";
CC  Nucleic Acids Res. 28:4317-4331(2000).
CC  -1- FUNCTION: Condenses 4-methyl-5-(beta-hydroxyethyl)-thiazole
CC  monophosphate (THZ-P) and 4-amino-5-hydroxymethyl pyrimidine
CC  pyrophosphate (HMP-PP) to form thiamine monophosphate (TMP) (By
CC  similarity).
CC  -1- CATALYTIC ACTIVITY: 2-methyl-4-amino-5-hydroxymethylpyrimidine
CC  diphosphate + 4-4-methyl-5-(2-phosphonooxyethyl)-thiazole -
CC  diphosphate + thiamine monophosphate.
CC  -1- PATHWAY: Thiamine biosynthesis.
CC  -1- SIMILARITY: BELONGS TO THE TMP-PPASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AP001512; BAB05150.1; -.
DR  HSSP; P39594; 2TPS.
DR  InterPro: IPR003733; TMP_synthase.
DR  pfam: PF02581; TMP-TENI.1.
DR  TIGRPFAM: TIGR00693; thtE.1.
KW  Thiamine biosynthesis; Transferase; Complete proteome.
SQ  SEQUENCE 211 AA; 23150 MW; 86AAB50608D2D6E CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 211;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  243 RVLQALA 249
DB  52 RVLQALA 58

RESULT 15
Y53L_SYNY3
ID  Y53L_SYNY3 STANDARD; PRT; 233 AA.
AC  P72583;
DT  15-JUL-1999 (Rel. 38, Created)
DT  15-JUL-1999 (Rel. 38, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Ycf53-like protein.
GN  SL0558.
OS  Synechocystis sp. (strain PCC 6803).
OC  Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX  NCBI_TaxID=1148;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=97061201; PubMed=8905231;
RA  Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA  Miyajima N., Hirosewa M., Sugliara M., Sasamoto S., Kimura T.,
RA  Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA  Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA  Tabata S.;
RT  "Sequence analysis of the genome of the unicellular cyanobacterium
RT  Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT  entire genome and assignment of potential protein-coding regions.";
RL  DNA Res. 3:109-136(1996).
CC  -1- SIMILARITY: BELONGS TO THE YCF53 FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D90899; BAA16582.1; -.
KW  Hypothetical protein; Complete proteome.

```

```

SQ  SEQUENCE 233 AA; 26465 MW; DE36B048AC83757C CRC64;
Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 233;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  968 LGSQDFE 974
DB  100 LGSQDFE 106

Search completed: May 7, 2003, 19:44:54
Job time : 19 secs

```



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:43:49 ; Search time 19 Seconds  
(without alignments)  
1593.484 Million cell updates/sec

Title: US-09-929-769-7

Perfect score: 1029

Sequence: 1 MHILVHAMVILLTGPPRA.....MDPSAQISEALRLHMEAVM 1029

Scoring table: OLIGO

Gapop 60.0 , Gapect 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUTS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.8	224	1 US-08-608-241-4	Sequence 4, Appl
2	8	0.8	224	2 US-08-922-182-4	Sequence 4, Appl
3	8	0.8	224	4 US-08-919-953-4	Sequence 4, Appl
4	8	0.8	226	4 US-09-192-983-4	Sequence 4, Appl
5	8	0.8	390	4 US-09-308-003-12	Sequence 12, Appl
6	8	0.8	396	4 US-09-134-001C-4580	Sequence 4580, Ap
7	7	0.7	27	2 US-08-660-789-9	Sequence 9, Appl
8	7	0.7	27	4 US-09-074-114-9	Sequence 9, Appl
9	7	0.7	30	1 US-08-127-351-8	Sequence 8, Appl
10	7	0.7	30	1 US-08-480-367B-8	Sequence 8, Appl
11	7	0.7	30	1 US-08-487-221A-8	Sequence 8, Appl
12	7	0.7	30	1 US-08-480-370-8	Sequence 8, Appl
13	7	0.7	33	1 US-08-127-351-9	Sequence 9, Appl
14	7	0.7	33	1 US-08-480-367B-9	Sequence 9, Appl
15	7	0.7	33	1 US-08-487-221A-9	Sequence 9, Appl
16	7	0.7	33	1 US-08-480-370-9	Sequence 9, Appl
17	7	0.7	35	4 US-09-460-145-9	Sequence 9, Appl
18	7	0.7	291	4 US-09-286-691-28	Sequence 28, Appl
19	7	0.7	291	4 US-09-687-147-28	Sequence 28, Appl
20	7	0.7	317	6 5340934-11	Patent No. 5340934
21	7	0.7	331	4 US-09-061-702-4	Sequence 4, Appl
22	7	0.7	331	4 US-09-069-023-32	Sequence 32, Appl
23	7	0.7	355	4 US-09-330-611-16	Sequence 16, Appl
24	7	0.7	395	3 US-08-981-825-6	Sequence 6, Appl
25	7	0.7	395	4 US-09-480-784-6	Sequence 6, Appl
26	7	0.7	446	2 US-08-874-138-6	Sequence 2, Appl
27	7	0.7	446	4 US-08-879-941-2	Sequence 2, Appl

28	7	0.7	446	4 US-09-747-116-2	Sequence 2, Appl
29	7	0.7	456	4 US-09-134-001C-2853	Sequence 2853, Ap
30	7	0.7	474	4 US-09-538-414-8	Sequence 8, Appl
31	7	0.7	572	2 US-09-032-315-7	Sequence 7, Appl
32	7	0.7	572	2 US-08-993-318A-7	Sequence 7, Appl
33	7	0.7	572	4 US-09-399-886-7	Sequence 7, Appl
34	7	0.7	572	4 US-09-396-260-7	Sequence 7, Appl
35	7	0.7	572	4 US-09-576-281-7	Sequence 7, Appl
36	7	0.7	573	4 US-09-042-709A-19	Sequence 19, Appl
37	7	0.7	574	1 US-08-140-729A-7	Sequence 7, Appl
38	7	0.7	574	1 US-08-546-666-7	Sequence 7, Appl
39	7	0.7	574	2 US-08-916-745-7	Sequence 4, Appl
40	7	0.7	574	2 US-08-948-569A-4	Sequence 4, Appl
41	7	0.7	574	2 US-08-663-808-6	Sequence 6, Appl
42	7	0.7	574	2 US-09-042-929-7	Sequence 7, Appl
43	7	0.7	574	2 US-08-546-661-7	Sequence 7, Appl
44	7	0.7	574	2 US-09-042-960-7	Sequence 7, Appl
45	7	0.7	574	2 US-09-188-469-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-08-608-241-4  
; Sequence 4, Application US/08608241  
; Patent No. 5747328  
; GENERAL INFORMATION:  
; APPLICANT: Donohue, Timothy J  
; APPLICANT: Barber, Robert D  
; APPLICANT: Witthuhn, Vernon  
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Plinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/608,241  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93511  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-9166  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 224 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-608-241-4

Query Match 0.8%; Score 8; DB 1; Length 224;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 PVVVSSSL 425  
|||||||  
DB 83 PVVVSSSL 90

RESULT 2  
US-08-922-182-4  
Sequence 4, Application US/08922182  
Patent No. 5834300  
GENERAL INFORMATION:  
APPLICANT: Donohue, Timothy J  
APPLICANT: Barber, Robert D  
APPLICANT: Witthuhn, Vernon  
TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE  
TITLE OF INVENTION: SENSING AND REMEDIATION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Plinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/922,182  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/608,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27,386  
REFERENCE/DOCKET NUMBER: 960296.93511  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-922-182-4  
Query Match 0.8%; Score 8; DB 2; Length 224;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 PVVVSSL 425  
Db 83 PVVVSSL 90  
RESULT 3  
US-08-919-953-4  
Sequence 4, Application US/08919953  
Patent No. 5837481  
GENERAL INFORMATION:  
APPLICANT: Donohue, Timothy J  
APPLICANT: Barber, Robert D  
APPLICANT: Witthuhn, Vernon  
TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE  
TITLE OF INVENTION: SENSING AND REMEDIATION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Plinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53703

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/919,953  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/608,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27,386  
REFERENCE/DOCKET NUMBER: 960296.93511  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-919-953-4  
Query Match 0.8%; Score 8; DB 2; Length 224;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 PVVVSSL 425  
Db 83 PVVVSSL 90  
RESULT 4  
US-09-192-983-4  
Sequence 4, Application US/09192983A  
Patent No. 6242244  
GENERAL INFORMATION:  
APPLICANT: Donohue, Timothy  
APPLICANT: Barber, Robert  
APPLICANT: Witthuhn, Vernon  
TITLE OF INVENTION: Microbial System for Formaldehyde Sensing and  
FILE REFERENCE: 960296.93505  
CURRENT APPLICATION NUMBER: US/09/192,983A  
CURRENT FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/919,953  
EARLIER FILING DATE: 1997-08-29  
EARLIER APPLICATION NUMBER: 08/608,241  
EARLIER FILING DATE: 1996-02-28  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 4  
LENGTH: 226  
TYPE: PRT  
ORGANISM: Rhododactyl sphaeroides  
US-09-192-983-4  
Query Match 0.8%; Score 8; DB 4; Length 226;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 PVVVSSL 425  
Db 85 PVVVSSL 92  
RESULT 5  
US-09-308-003-12  
Sequence 12, Application US/09308003

Patent No. 6326170  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,  
FILE REFERENCE: GMI10093  
CURRENT APPLICATION NUMBER: US/09/308,003  
CURRENT FILING DATE: 1999-05-10  
EARLIER APPLICATION NUMBER: 60/058,710  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 390  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-308-003-12

Query Match  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 870 LPLVSVSL 877  
|||||||  
Db 312 LPLVSVSL 319

RESULT 6  
US-09-134-001C-4580  
Sequence 4580, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4580  
LENGTH: 396  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4580

Query Match  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 870 LPLVSVSL 877  
|||||||  
Db 317 LPLVSVSL 324

RESULT 7  
US-08-660-789-9  
Sequence 9, Application US/08660789  
Patent No. 5843405  
GENERAL INFORMATION:  
APPLICANT: Middeldorp, Jaap M.  
TITLE OF INVENTION: Epstein Barr Virus peptides and  
TITLE OF INVENTION: antibodies against these peptides  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo No. 5843405el Patent Department  
STREET: 1300 Piccard Drive, Suite 206  
CITY: Rockville

STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,789  
FILING DATE: 06-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gormley, Mary E.  
REGISTRATION NUMBER: 34,409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
TELEFAX: (301) 977-0847  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Epstein-Barr virus  
US-08-660-789-9

Query Match  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 324 ASQASAG 330  
|||||||  
Db 15 ASQASAG 21

RESULT 8  
US-09-074-114-9  
Sequence 9, Application US/09074114  
Patent No. 6143865  
GENERAL INFORMATION:  
APPLICANT: Middeldorp, Jaap Michel  
TITLE OF INVENTION: Epstein Barr Virus peptides and  
TITLE OF INVENTION: antibodies against these peptides  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo No. 6143865el Patent Dept.  
STREET: 1300 Piccard Drive, Suite 206  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,114  
FILING DATE: 07-MAY-1998  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Klesner, Sharon N.  
REGISTRATION NUMBER: 36,335  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-948-7400  
TELEFAX: 301-948-9751  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Epstein-Barr virus  
US-09-074-114-9

Query Match 0.7%; Score 7; DB 4; Length 27;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 ASOASAG 330  
Db 15 ASOASAG 21

RESULT 9  
US-08-127-351-8  
Sequence 8, Application US/08127351  
Patent No. 5449761

GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/127,351  
FILING DATE: 28-SEP-1993  
CLASSIFICATION: 534  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 4980-004-44  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide

US-08-127-351-8

Query Match 0.7%; Score 7; DB 1; Length 30;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LVDALQ 76  
Db 10 LVDALQ 16

RESULT 10  
US-08-480-367B-8  
Sequence 8, Application US/08480367B

Patent No. 5578288  
GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER  
STREET: 99 Canal Center Plaza, Suite 300  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22314

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,367B  
FILING DATE: 07-06-95  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 2654-002A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 684-1111  
TELEFAX: (703) 684-1124  
TELEX:

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide

US-08-480-367B-8

Query Match 0.7%; Score 7; DB 1; Length 30;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LVDALQ 76  
Db 10 LVDALQ 16

RESULT 11  
US-08-487-221A-8  
Sequence 8, Application US/08487221A  
Patent No. 5593656  
GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,221A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/127,351  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 4980-004-44  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-487-221A-8

Query Match 0.7%; Score 7; DB 1; Length 30;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LVDAALQ 76  
|||||||  
DB 10 LVDAALQ 16

RESULT 12  
US-08-480-370-8  
Sequence 8, Application US/08480370  
Patent No. 5609847  
GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,370  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/127,351  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 4980-004-44  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-480-370-8

Query Match 0.7%; Score 7; DB 1; Length 30;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LVDAALQ 76  
|||||||  
DB 10 LVDAALQ 16

RESULT 13  
US-08-127-351-9  
Sequence 9, Application US/08127351  
Patent No. 5449761  
GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/127,351  
FILING DATE: 28-SEP-1993  
CLASSIFICATION: 534  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 4980-004-44  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-127-351-9

Query Match 0.7%; Score 7; DB 1; Length 33;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LVDAALQ 76  
|||||||  
DB 10 LVDAALQ 16

RESULT 14  
US-08-480-367B-9  
Sequence 9, Application US/08480367B  
Patent No. 5578288  
GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER  
STREET: 99 Canal Center Plaza, Suite 300  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,367B  
FILING DATE: 07-06-95  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 2654-002A  
TELEPHONE: (703) 684-1111  
TELEFAX: (703) 684-1124  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-480-367B-9

Query Match 0.7%; Score 7; DB 1; Length 33;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LVDAALQ 76  
Db 10 LVDAALQ 16

RESULT 15  
US-08-487-221A-9  
Sequence 9, Application US/08487221A  
Patent No. 5593656  
GENERAL INFORMATION:  
APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.  
APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER &  
ADDRESSEE: NEUSTADY, P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.

ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,221A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/127,351  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 4980-004-44  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-487-221A-9

Query Match 0.7%; Score 7; DB 1; Length 33;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LVDAALQ 76  
Db 10 LVDAALQ 16

Search completed: May 7, 2003, 19:46:46  
Job time : 22 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:44:39 ; Search time 29 Seconds  
(Without alignments)  
3265.326 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 1029  
Sequence: 1 MHILVHVAMVILTLIGPPRA.....MDPSAQISEALRIILHMEAVM 1029

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size: 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

Published Applications\_AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep: \*  
2: /cgn2\_6/ptodata/2/pubpaa/PCIT\_NEW\_PUB pep: \*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep: \*  
4: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep: \*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep: \*  
6: /cgn2\_6/ptodata/2/pubpaa/PCITUS\_PUBCOMB pep: \*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep: \*  
8: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep: \*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep: \*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep: \*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep: \*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep: \*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep: \*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1029	100.0	1029	9	US-10-033-245-22
2	1029	100.0	1029	9	US-10-033-223-22
3	1029	100.0	1029	9	US-10-033-167-22
4	1029	100.0	1029	9	US-10-033-244-22
5	1029	100.0	1029	9	US-10-033-435-22
6	1029	100.0	1029	9	US-10-033-990-22
7	1029	100.0	1029	9	US-09-929-769-7
8	1029	100.0	1029	9	US-10-033-996-22
9	1029	100.0	1029	12	US-10-033-396-22
10	1029	100.0	1029	12	US-10-033-246-22
11	1029	100.0	1029	12	US-10-033-301-22
12	1029	100.0	1029	12	US-10-033-326-22
13	404	39.3	404	9	US-09-984-271-235
14	327	31.8	328	9	US-09-984-271-148
15	43	4.2	161	9	US-09-764-891-4155
16	8	0.8	209	10	US-09-893-737-126
17	8	0.8	253	9	US-09-774-381-54
18	7	0.7	9	9	US-09-922-226-88
19	7	0.7	27	10	US-09-864-761-46828

20	7	0.7	54	10	US-09-864-761-46415	Sequence 46415, A
21	7	0.7	60	10	US-09-864-761-37408	Sequence 37408, A
22	7	0.7	62	9	US-09-764-868-1132	Sequence 1132, Ap
23	7	0.7	73	9	US-09-981-876-139	Sequence 139, App
24	7	0.7	73	9	US-09-148-545-139	Sequence 139, App
25	7	0.7	95	9	US-10-016-6344-125	Sequence 125, App
26	7	0.7	99	10	US-09-864-761-46653	Sequence 46653, A
27	7	0.7	116	10	US-09-864-761-44770	Sequence 44770, A
28	7	0.7	172	9	US-09-847-208-50	Sequence 50, Appl
29	7	0.7	181	9	US-10-163-499-14	Sequence 14, Appl
30	7	0.7	181	9	US-10-163-499-16	Sequence 16, Appl
31	7	0.7	197	9	US-10-097-065-135	Sequence 135, App
32	7	0.7	208	10	US-09-214-881A-11	Sequence 11, Appl
33	7	0.7	233	10	US-09-938-803-15	Sequence 15, Appl
34	7	0.7	262	10	US-09-960-738A-2	Sequence 2, Appl
35	7	0.7	271	10	US-09-846-808-20	Sequence 20, Appl
36	7	0.7	276	9	US-09-738-626-5599	Sequence 5599, Ap
37	7	0.7	281	9	US-10-101-464A-518	Sequence 518, App
38	7	0.7	283	9	US-09-895-913A-28	Sequence 28, Appl
39	7	0.7	283	10	US-09-815-242-14063	Sequence 14063, A
40	7	0.7	286	9	US-09-997-816-2	Sequence 2, Appl
41	7	0.7	298	10	US-09-815-242-5142	Sequence 5142, Ap
42	7	0.7	315	10	US-09-741-669-361	Sequence 361, App
43	7	0.7	319	9	US-10-116-722A-6	Sequence 6, Appl
44	7	0.7	319	9	US-10-116-722A-8	Sequence 8, Appl
45	7	0.7	320	10	US-09-741-669-467	Sequence 467, App

## ALIGNMENTS

RESULT 1  
US-10-033-245-22  
Sequence 22, Application US/10033245  
Patent No. US20020160392A1  
GENERAL INFORMATION:  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Gunney, Austin L.  
APPLICANT: Pany, James  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2930R1C7  
CURRENT APPLICATION NUMBER: US/10/033,245  
CURRENT FILING DATE: 2001-12-27  
PRIOR APPLICATION NUMBER: 60/095,325  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/112,851  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,145  
PRIOR FILING DATE: 1998-12-16  
PRIOR APPLICATION NUMBER: 60/113,511  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/115,558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,565  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115,733  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119,341  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: 60/119,537  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/119,965

PRIOR FILING DATE: 1999-02-12  
 PRIOR APPLICATION NUMBER: 60/162,506  
 PRIOR FILING DATE: 1999-10-29  
 PRIOR APPLICATION NUMBER: 60/170,262  
 PRIOR FILING DATE: 1999-12-09  
 PRIOR APPLICATION NUMBER: 60/187,202  
 PRIOR FILING DATE: 2000-03-03  
 PRIOR APPLICATION NUMBER: PCT/US99/12252  
 PRIOR FILING DATE: 1999-06-02  
 PRIOR APPLICATION NUMBER: PCT/US99/28634  
 PRIOR FILING DATE: 1999-12-01  
 PRIOR APPLICATION NUMBER: PCT/US99/28551  
 PRIOR FILING DATE: 1999-12-02  
 PRIOR APPLICATION NUMBER: PCT/US00/03565  
 PRIOR FILING DATE: 2000-02-11  
 PRIOR APPLICATION NUMBER: PCT/US00/04414  
 PRIOR FILING DATE: 2000-02-22  
 PRIOR APPLICATION NUMBER: PCT/US00/05841  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: PCT/US00/08439  
 PRIOR FILING DATE: 2000-03-30  
 PRIOR APPLICATION NUMBER: PCT/US00/14941  
 PRIOR FILING DATE: 2000-05-30  
 PRIOR APPLICATION NUMBER: PCT/US00/15264  
 PRIOR FILING DATE: 2000-06-02  
 PRIOR APPLICATION NUMBER: PCT/US00/32678  
 PRIOR FILING DATE: 2000-12-01  
 NUMBER OF SEQ ID NOS: 38  
 SEQ ID NO 22  
 LENGTH: 1029  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-033-245-22

Query Match 100.0%; Score 1029; DB 9; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHTLVNAMYLTLLGPRRDSDFQALLDIWPEEKPLPAFLVDTSEALLPWLKL 60  
 DB 1 MHTLVNAMYLTLLGPRRDSDFQALLDIWPEEKPLPAFLVDTSEALLPWLKL 60  
 QY 61 RMIRSEVLRVDAADLEPOQLLFVQSGFIPVSSMSKLLQFDQAVADHPOTLEONIM 120  
 DB 61 RMIRSEVLRVDAADLEPOQLLFVQSGFIPVSSMSKLLQFDQAVADHPOTLEONIM 120  
 QY 121 DKNYMAHLVEVQHERGASGGOTFHSLLTASLPERRDSTEAPKPSSEOPIGGRIRVGT 180  
 DB 121 DKNYMAHLVEVQHERGASGGOTFHSLLTASLPERRDSTEAPKPSSEOPIGGRIRVGT 180  
 QY 181 QLRVLGEDDLACGFLQIFPLSPRRWSSPPRVALLAQQALQGLARVYVQSGPEVPGI 240  
 DB 181 QLRVLGEDDLACGFLQIFPLSPRRWSSPPRVALLAQQALQGLARVYVQSGPEVPGI 240  
 QY 241 TVRYLQALATLLSSPHGALVMSMHRSHFLACPLLRQLCYORCPVODTGGSSLEFLVLL 300  
 DB 241 TVRYLQALATLLSSPHGALVMSMHRSHFLACPLLRQLCYORCPVODTGGSSLEFLVLL 300  
 QY 301 QMLQWLDSPEVEGGPLRAQDRLMLASQASAGRRLSDVNGGLRLRLAELAFRODLEVVSTV 360  
 DB 301 QMLQWLDSPEVEGGPLRAQDRLMLASQASAGRRLSDVNGGLRLRLAELAFRODLEVVSTV 360  
 QY 361 RAVIATIRSEQCSVEPDLISKVQLGILEVRSPHLEELLTAFFSATADAASPPACKPVV 420  
 DB 361 RAVIATIRSEQCSVEPDLISKVQLGILEVRSPHLEELLTAFFSATADAASPPACKPVV 420  
 QY 421 VVSSLLOEEEPGLAGGRGADGSLFAVRGLGSSGLLVMDLEMDPEVVSSCPDLQRL 480  
 DB 421 VVSSLLOEEEPGLAGGRGADGSLFAVRGLGSSGLLVMDLEMDPEVVSSCPDLQRL 480  
 QY 481 FSRKRGKQAVPSFRYLLTLFTTHOSSMPTLHQCIRVLLGKRSREGFDDSSASLDPLMAC 540  
 DB 481 FSRKRGKQAVPSFRYLLTLFTTHOSSMPTLHQCIRVLLGKRSREGFDDSSASLDPLMAC 540

QY 541 IHVRIMQGRDQRTQKRRRELYLRVQGPGLISLVELILAEATRSODGDTAACSLIOAR 600  
 DB 541 IHVRIMQGRDQRTQKRRRELYLRVQGPGLISLVELILAEATRSODGDTAACSLIOAR 600  
 QY 601 LPLLSCCGDDESVKRVTEHLSCIOQKGDVSGRCRDLLOLYQREPRLPVPEVL 660  
 DB 601 LPLLSCCGDDESVKRVTEHLSCIOQKGDVSGRCRDLLOLYQREPRLPVPEVL 660  
 QY 661 LHSEGAASSVCKLDGLIHRTTLADTSDSRALENGADASMACRLAAVAPLLRL 720  
 DB 661 LHSEGAASSVCKLDGLIHRTTLADTSDSRALENGADASMACRLAAVAPLLRL 720  
 QY 721 PMIALHGRTHLNFQEFROONHLSCLHVLGLLELLQPHVFRSEHOGALMDCLSLFRL 780  
 DB 721 PMIALHGRTHLNFQEFROONHLSCLHVLGLLELLQPHVFRSEHOGALMDCLSLFRL 780  
 QY 781 LLNVRKSSRHAAFINKEVQFIHKYIYVNAAPATSFLOKADPDLHDSFNSDLYMKSL 840  
 DB 781 LLNVRKSSRHAAFINKEVQFIHKYIYVNAAPATSFLOKADPDLHDSFNSDLYMKSL 840  
 QY 841 LAGLSLRDNRDRLGDEEGEESAGSLPLVSVSLFTPLTAEMAPYKRLRSGOTVE 900  
 DB 841 LAGLSLRDNRDRLGDEEGEESAGSLPLVSVSLFTPLTAEMAPYKRLRSGOTVE 900  
 QY 901 DLEVLSDIDEMSRRELLSFSTNQLRLMSSAECCRLAASLARSQNSPSIAAF 960  
 DB 901 DLEVLSDIDEMSRRELLSFSTNQLRLMSSAECCRLAASLARSQNSPSIAAF 960  
 QY 961 LPTMYCLGSODPEVOTALRNLPYVALLOEHAVALHRAFLVGMVGMQNDPSAQISEAL 1020  
 DB 961 LPTMYCLGSODPEVOTALRNLPYVALLOEHAVALHRAFLVGMVGMQNDPSAQISEAL 1020  
 QY 1021 RLHMEAVM 1029  
 DB 1021 RLHMEAVM 1029

RESULT 2  
 US-10-033-223-22  
 ; Sequence 22, Application US/10033223  
 ; Patent No. US20020164646A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Botstein, David  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Ferrara, Napoleone  
 ; APPLICANT: Fong, Sherman  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Roy, Margaret Ann  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 ; FILE REFERENCE: P2930R1C9  
 ; CURRENT APPLICATION NUMBER: US/10/033,223  
 ; PRIOR FILING DATE: 2001-12-27  
 ; PRIOR APPLICATION NUMBER: 60/095,325  
 ; PRIOR FILING DATE: 1998-08-04  
 ; PRIOR APPLICATION NUMBER: 60/112,851  
 ; PRIOR FILING DATE: 1998-12-16  
 ; PRIOR APPLICATION NUMBER: 60/113,145  
 ; PRIOR FILING DATE: 1998-12-16  
 ; PRIOR APPLICATION NUMBER: 60/113,511  
 ; PRIOR FILING DATE: 1998-12-22  
 ; PRIOR APPLICATION NUMBER: 60/115,558  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/115,565  
 ; PRIOR FILING DATE: 1999-01-12



```

: PRIOR APPLICATION NUMBER: 60/115,733
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/119,341
: PRIOR FILING DATE: 1999-02-09
: PRIOR APPLICATION NUMBER: 60/119,537
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: 60/119,965
: PRIOR FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: 60/162,506
: PRIOR FILING DATE: 1999-10-29
: PRIOR APPLICATION NUMBER: 60/170,262
: PRIOR FILING DATE: 1999-12-09
: PRIOR APPLICATION NUMBER: 60/187,202
: PRIOR FILING DATE: 2000-03-03
: PRIOR APPLICATION NUMBER: PCT/US99/12252
: PRIOR FILING DATE: 1999-06-02
: PRIOR APPLICATION NUMBER: PCT/US99/28634
: PRIOR FILING DATE: 1999-12-01
: PRIOR APPLICATION NUMBER: PCT/US99/28551
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US00/03565
: PRIOR FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: PCT/US00/05841
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: PCT/US00/08439
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: PCT/US00/14941
: PRIOR FILING DATE: 2000-05-30
: PRIOR APPLICATION NUMBER: PCT/US00/15264
: PRIOR FILING DATE: 2000-06-02
: PRIOR APPLICATION NUMBER: PCT/US00/32678
: PRIOR FILING DATE: 2000-12-01
: NUMBER OF SEQ ID NOS: 38
: SEQ ID NO 22
: LENGTH: 1029
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-033-223-22

```

```

Query Match      100.0%; Score 1029; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MHILVHAMVILTLGPPRADSEFOALDIFPEEKPLPTAFIVDTSEALLPDWKL 60
DB 1 MHILVHAMVILTLGPPRADSEFOALDIFPEEKPLPTAFIVDTSEALLPDWKL 60
QY 61 RMIRSEVLRLVDALQDLEPOQLLEFVOSFGIPVSSMSKILQFLDOAVAHDPQILEQNI 120
DB 61 RMIRSEVLRLVDALQDLEPOQLLEFVOSFGIPVSSMSKILQFLDOAVAHDPQILEQNI 120
QY 121 DKNYMAHLVEVOHERGASGGQTFHSHLTLASLPRRDSTEARPKSSPEOPIQGRIRVGT 180
DB 121 DKNYMAHLVEVOHERGASGGQTFHSHLTLASLPRRDSTEARPKSSPEOPIQGRIRVGT 180
QY 181 QLRVIGPDDDLAGMFLQIFPLSPDRMOSSSPRPAVALQALGELAVVVGSEVPPI 240
DB 181 QLRVIGPDDDLAGMFLQIFPLSPDRMOSSSPRPAVALQALGELAVVVGSEVPPI 240
QY 241 TVRVLOALATLTLSSPHGALVMSHRSHFLACPLRLQOLCOYRCVPQDTEGSESLFLKVL 300
DB 241 TVRVLOALATLTLSSPHGALVMSHRSHFLACPLRLQOLCOYRCVPQDTEGSESLFLKVL 300
QY 301 QMLQMLDSPGVGGPLRLQRLMLASQASAGRRLSVRGGLRLALALAFRODLEVVSTIV 360
DB 301 QMLQMLDSPGVGGPLRLQRLMLASQASAGRRLSVRGGLRLALALAFRODLEVVSTIV 360
QY 361 RAVITLTSGECCSVPEPLDISKVLQGLLEVRSPHLELLTAFESATADASFPACKPV 420
DB 361 RAVITLTSGECCSVPEPLDISKVLQGLLEVRSPHLELLTAFESATADASFPACKPV 420

```

```

QY 421 VVSSILLQEEEPFLAGKPGADGSLFVAVRLGPSSGLVDMLEMLDPEVYSSCPDQLRL 480
DB 421 VVSSILLQEEEPFLAGKPGADGSLFVAVRLGPSSGLVDMLEMLDPEVYSSCPDQLRL 480
QY 481 FSRKRGKGOAVPSFRPYLLTLFTHQSSWPTLHOCTIRVLLGKSRQOREPDSALDFLWAC 540
DB 481 FSRKRGKGOAVPSFRPYLLTLFTHQSSWPTLHOCTIRVLLGKSRQOREPDSALDFLWAC 540
QY 541 HVPRHWGRDORTQKRRREELVAVQGPILLSIVELLIAEFTSQSDDTAACSILQAR 600
DB 541 HVPRHWGRDORTQKRRREELVAVQGPILLSIVELLIAEFTSQSDDTAACSILQAR 600
QY 601 LPILLSGCCGDESVKRYTEHLSGCIQMGDSVLGRGRCDLLQLYLQRPETARVPEYL 660
DB 601 LPILLSGCCGDESVKRYTEHLSGCIQMGDSVLGRGRCDLLQLYLQRPETARVPEYL 660
QY 661 LHSEGAASSVCKLDGLHRTITLLADTSDSRALENRGADASMACRKLAVAHPLLLRL 720
DB 661 LHSEGAASSVCKLDGLHRTITLLADTSDSRALENRGADASMACRKLAVAHPLLLRL 720
QY 721 PMIAALHGRTHLNFQERQONHLSCFLHVLGLLELQPHVRSEHOGALMPCILLSFIRL 780
DB 721 PMIAALHGRTHLNFQERQONHLSCFLHVLGLLELQPHVRSEHOGALMPCILLSFIRL 780
QY 781 LLYNRKSSRHIAAFINKFVQFIHKYITTYNAPALISFLQKHADPLHDLSFDSNDLYWLKSL 840
DB 781 LLYNRKSSRHIAAFINKFVQFIHKYITTYNAPALISFLQKHADPLHDLSFDSNDLYWLKSL 840
QY 841 LAGLSLPSRDRDTRDGLDEEGEESAGSLPVYSLSFTPLTAEMAPYMKRLSKQIVE 900
DB 841 LAGLSLPSRDRDTRDGLDEEGEESAGSLPVYSLSFTPLTAEMAPYMKRLSKQIVE 900
QY 901 DLLEVLSDIDEMSRREPILTSFSTMLQRLMSAECCCNLAFLSALRSMQSPSTAAAF 960
DB 901 DLLEVLSDIDEMSRREPILTSFSTMLQRLMSAECCCNLAFLSALRSMQSPSTAAAF 960
QY 961 LPTFMVCLGSDPEFVYQATRLNLPETAYALCOEHAVALHRAFLVGYGMDPSAQISEAL 1020
DB 961 LPTFMVCLGSDPEFVYQATRLNLPETAYALCOEHAVALHRAFLVGYGMDPSAQISEAL 1020
QY 1021 RILHMEAVM 1029
DB 1021 RILHMEAVM 1029

```

```

RESULT 3
US-10-033-167-22
: Sequence 22, Application US/10033167
: Publication No. US20020182618A1
: GENERAL INFORMATION:
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Goddard, Audrey
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: FILE REFERENCE: P2930RC10
: CURRENT FILING DATE: 2001-12-27
: PRIOR APPLICATION NUMBER: US/10/033,167
: PRIOR FILING DATE: 2001-12-27
: PRIOR APPLICATION NUMBER: 60/095,325
: PRIOR FILING DATE: 1998-08-04
: PRIOR APPLICATION NUMBER: 60/112,851
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,145

```

PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,511  
 PRIOR FILING DATE: 1998-12-22  
 PRIOR APPLICATION NUMBER: 60/115,558  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,565  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,733  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/119,341  
 PRIOR FILING DATE: 1999-02-09  
 PRIOR APPLICATION NUMBER: 60/119,537  
 PRIOR FILING DATE: 1999-02-10  
 PRIOR APPLICATION NUMBER: 60/119,965  
 PRIOR FILING DATE: 1999-02-12  
 PRIOR APPLICATION NUMBER: 60/162,506  
 PRIOR FILING DATE: 1999-10-29  
 PRIOR APPLICATION NUMBER: 60/170,262  
 PRIOR FILING DATE: 1999-12-09  
 PRIOR APPLICATION NUMBER: 60/187,202  
 PRIOR FILING DATE: 2000-03-03  
 PRIOR APPLICATION NUMBER: PCT/US99/12252  
 PRIOR FILING DATE: 1999-06-02  
 PRIOR APPLICATION NUMBER: PCT/US99/28634  
 PRIOR FILING DATE: 1999-12-01  
 PRIOR APPLICATION NUMBER: PCT/US99/28551  
 PRIOR FILING DATE: 1999-12-02  
 PRIOR APPLICATION NUMBER: PCT/US00/03565  
 PRIOR FILING DATE: 2000-02-11  
 PRIOR APPLICATION NUMBER: PCT/US00/04414  
 PRIOR FILING DATE: 2000-02-22  
 PRIOR APPLICATION NUMBER: PCT/US00/05841  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: PCT/US00/08439  
 PRIOR FILING DATE: 2000-03-30  
 PRIOR APPLICATION NUMBER: PCT/US00/14941  
 PRIOR FILING DATE: 2000-05-30  
 PRIOR APPLICATION NUMBER: PCT/US00/15264  
 PRIOR FILING DATE: 2000-06-02  
 PRIOR APPLICATION NUMBER: PCT/US00/32678  
 PRIOR FILING DATE: 2000-12-01  
 NUMBER OF SEQ ID NOS: 38  
 SEQ ID NO 22  
 LENGTH: 1029  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-033-167-22

Query Match 100.0% Score 1029 DB 9 Length 1029  
 Best Local Similarity 100.0% Pred. No. 0  
 Matches 1029: Conservative 0; Mismatches 0; Indels 0; Gaps 0  
 1 MHLLVNAWYLLTLGPRADSEFOALDIWPEEKRLPTAFVDTSEBLLPWLK 60  
 1 MHLLVNAWYLLTLGPRADSEFOALDIWPEEKRLPTAFVDTSEBLLPWLK 60  
 61 RMIRSEYLRVDAADLDLEQQLLFYQSGIYVSSKSLLOFLDQAVAHDPOTLEONIM 120  
 61 RMIRSEYLRVDAADLDLEQQLLFYQSGIYVSSKSLLOFLDQAVAHDPOTLEONIM 120  
 121 DKNYMAHLYVOHERGASGOTFHSLLTASLPPRDSLEAPKPKSSPEOPIGGGRIVGT 180  
 121 DKNYMAHLYVOHERGASGOTFHSLLTASLPPRDSLEAPKPKSSPEOPIGGGRIVGT 180  
 121 DKNYMAHLYVOHERGASGOTFHSLLTASLPPRDSLEAPKPKSSPEOPIGGGRIVGT 180  
 181 QLRVLGPEDDLAGMFLQIFPLSPDRKMOSSPRVALALQALGOLARVVOGSPFVGI 240  
 181 QLRVLGPEDDLAGMFLQIFPLSPDRKMOSSPRVALALQALGOLARVVOGSPFVGI 240  
 241 TVRYLQALATLSSPHGALVMSHNSHFLACPLRLQCOYRCVPQDTGFSFLKVL 300  
 241 TVRYLQALATLSSPHGALVMSHNSHFLACPLRLQCOYRCVPQDTGFSFLKVL 300  
 301 QMLQWLDSPVEGGLPLRAQLMLASQASAGRRLSDVNGGLRLAELAFRODLEVVSSV 360

301 QMLQWLDSPVEGGLPLRAQLMLASQASAGRRLSDVNGGLRLAELAFRODLEVVSSV 360  
 361 RAVIATLRSEGCQSVFEDLISKVLOGLIEVRSFHEBELLAFFSATADAAAPPACKPVY 420  
 361 RAVIATLRSEGCQSVFEDLISKVLOGLIEVRSFHEBELLAFFSATADAAAPPACKPVY 420  
 421 VVSSLIQEEEPPLAGCKRGADGSLAVRLGPPSGGLVMDLEMDPEVSSCPDQLRL 480  
 421 VVSSLIQEEEPPLAGCKRGADGSLAVRLGPPSGGLVMDLEMDPEVSSCPDQLRL 480  
 481 FSRKKGQAQVPSFRYLLTLFTHOSWPTLHOCIRVLLGKSREOFDPSASIDFLMAC 540  
 481 FSRKKGQAQVPSFRYLLTLFTHOSWPTLHOCIRVLLGKSREOFDPSASIDFLMAC 540  
 541 IHVRIWQGRDQRTPOKREELVRYOGPELISVELLAEAEFRSDGDTAACSLIOAR 600  
 541 IHVRIWQGRDQRTPOKREELVRYOGPELISVELLAEAEFRSDGDTAACSLIOAR 600  
 601 LPLILSCCGDDSVKRYTEHLSGCTIOQWGDVYGRRCRLQLQYLOPELRYVPEVL 660  
 601 LPLILSCCGDDSVKRYTEHLSGCTIOQWGDVYGRRCRLQLQYLOPELRYVPEVL 660  
 661 LHSBGAASSVCKLDGLIHRFTLLADTSDSRALENRGADAMACRKLAVAHPLLLRL 720  
 661 LHSBGAASSVCKLDGLIHRFTLLADTSDSRALENRGADAMACRKLAVAHPLLLRL 720  
 721 PMIAALLHGRTHLNFOEFRQOONHLSCLFHLVGLLELLOPHVFRSEHOGALMDCLSFIRL 780  
 721 PMIAALLHGRTHLNFOEFRQOONHLSCLFHLVGLLELLOPHVFRSEHOGALMDCLSFIRL 780  
 781 LMYRKSSRLAFLINKFVFIHXYTYNAPPAISFLQKADPLHDLSFQNSDLVMLKSL 840  
 781 LMYRKSSRLAFLINKFVFIHXYTYNAPPAISFLQKADPLHDLSFQNSDLVMLKSL 840  
 841 LAGLSLPSRDRDRGDEGEESAGSLPLVSVSLFPTLTAEMAPYKRLSRGOTVE 900  
 841 LAGLSLPSRDRDRGDEGEESAGSLPLVSVSLFPTLTAEMAPYKRLSRGOTVE 900  
 901 DLLEVLSDIDEMSRREILSFSTNLQRLMSSAECCRLAFSLALRSMONSPSIAAF 960  
 901 DLLEVLSDIDEMSRREILSFSTNLQRLMSSAECCRLAFSLALRSMONSPSIAAF 960  
 961 LPTFMVCLGSDDEEVOTALRNLPEYALLCOEHAAVLLHRAFLVGMGMDSOISSEAL 1020  
 961 LPTFMVCLGSDDEEVOTALRNLPEYALLCOEHAAVLLHRAFLVGMGMDSOISSEAL 1020  
 1021 RILHMEAVM 1029  
 1021 RILHMEAVM 1029

RESULT 4  
 US-10-033-244-22  
 Publication No. US20020192668A1  
 GENERAL INFORMATION:  
 APPLICANT: Bolstein, David  
 APPLICANT: Desnoyers, Luc  
 APPLICANT: Ferrara, Napoleone  
 APPLICANT: Fong, Sherman  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Pan, James  
 APPLICANT: Roy, Margaret Ann  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumas, Daniel  
 APPLICANT: Watanabe, Colin K.  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 TITLE OF INVENTION: Acids Encoding the Same  
 FILE REFERENCE: P2930R1C2

```

: CURRENT APPLICATION NUMBER: US/10/033,244
: CURRENT FILING DATE: 2001-12-27
: PRIOR APPLICATION NUMBER: 60/095,325
: PRIOR FILING DATE: 1998-08-04
: PRIOR APPLICATION NUMBER: 60/112,851
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,145
: PRIOR FILING DATE: 1998-12-16
: PRIOR APPLICATION NUMBER: 60/113,511
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: 60/115,558
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/115,565
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/115,733
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/119,341
: PRIOR FILING DATE: 1999-02-09
: PRIOR APPLICATION NUMBER: 60/119,537
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: 60/119,965
: PRIOR FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: 60/162,506
: PRIOR FILING DATE: 1999-10-29
: PRIOR APPLICATION NUMBER: 60/170,262
: PRIOR FILING DATE: 1999-12-09
: PRIOR APPLICATION NUMBER: 60/187,202
: PRIOR FILING DATE: 2000-03-03
: PRIOR APPLICATION NUMBER: PCT/US99/12252
: PRIOR FILING DATE: 1999-06-02
: PRIOR APPLICATION NUMBER: PCT/US99/28634
: PRIOR FILING DATE: 1999-12-01
: PRIOR APPLICATION NUMBER: PCT/US99/28551
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US00/03565
: PRIOR FILING DATE: 2000-02-11
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: PCT/US00/05841
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: PCT/US00/08439
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: PCT/US00/14941
: PRIOR FILING DATE: 2000-05-30
: PRIOR APPLICATION NUMBER: PCT/US00/15264
: PRIOR FILING DATE: 2000-06-02
: PRIOR APPLICATION NUMBER: PCT/US00/32678
: PRIOR FILING DATE: 2000-12-01
: NUMBER OF SEQ ID NOS: 38
: SEQ ID NO 22
: LENGTH: 1029
: TYPE: prt
: ORGANISM: Homo sapiens
US-10-033-244-22

Query Match          100.0%; Score 1029; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No.: 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 181 QLRVLGPEDDLGMFLQIFPLSPDPRWSSSPRVALALQALGQELARVVOGSEVPVGI 240
QY 241 TVRVLAATLTLSSPHGALVMSMRSHFLACPLRLQLOYORCVODTGFSSFLKVL 300
Db 241 TVRVLAATLTLSSPHGALVMSMRSHFLACPLRLQLOYORCVODTGFSSFLKVL 300
QY 301 QMLQWIDSPGVEGCPRLAOLRLMASQASAGRRLSDVRGGLRLAEALAFRODLEVSVTV 360
Db 301 QMLQWIDSPGVEGCPRLAOLRLMASQASAGRRLSDVRGGLRLAEALAFRODLEVSVTV 360
QY 361 RAVIATLRSGECCSVEPDLISKVLOGLIEVRSPhIEELLTAFFSATADAAPPACKPVV 420
Db 361 RAVIATLRSGECCSVEPDLISKVLOGLIEVRSPhIEELLTAFFSATADAAPPACKPVV 420
QY 421 VVSILLQEEEPPLAGKPPADGSLAVPLGSSGLVLMLELDEEVVSSCPDLRL 480
Db 421 VVSILLQEEEPPLAGKPPADGSLAVPLGSSGLVLMLELDEEVVSSCPDLRL 480
QY 481 FSRKKGQAOYPSFRPYLLTFTHSSWPTLHQCIRVLGKSREQRPDSASLDFLMAC 540
Db 481 FSRKKGQAOYPSFRPYLLTFTHSSWPTLHQCIRVLGKSREQRPDSASLDFLMAC 540
QY 541 IHVPRIWGRDORTPOKRREELVLRVQPELISVLELLIAEFTNSQDGDITACSLQAR 600
Db 541 IHVPRIWGRDORTPOKRREELVLRVQPELISVLELLIAEFTNSQDGDITACSLQAR 600
QY 601 LPLLSCCGDDESVKRVTEHLSGCTIOQWGDVLRGCRDILLQLOLORPELRVPPEVL 660
Db 601 LPLLSCCGDDESVKRVTEHLSGCTIOQWGDVLRGCRDILLQLOLORPELRVPPEVL 660
QY 661 LHSEGAASSVCKLQGLIHRFTLLADTSDSRALENRGADASMACRKLAVAPHLRL 720
Db 661 LHSEGAASSVCKLQGLIHRFTLLADTSDSRALENRGADASMACRKLAVAPHLRL 720
QY 721 PMIAALHRTHLNROEFQOQNLSCFLVYLGIELLOHVHRSSEQALMDCLSFTL 780
Db 721 PMIAALHRTHLNROEFQOQNLSCFLVYLGIELLOHVHRSSEQALMDCLSFTL 780
QY 781 LNYRKSSSHLLAFINKEFOFIHKYTYNAPAAISFLQKHADPHLSPDNSDLVLMKSL 840
Db 781 LNYRKSSSHLLAFINKEFOFIHKYTYNAPAAISFLQKHADPHLSPDNSDLVLMKSL 840
QY 841 LAGLSLPSRDDRTDRLGDEGEESAGSLPLVSVSLFTPLAAEAPYMKRLSRQOVE 900
Db 841 LAGLSLPSRDDRTDRLGDEGEESAGSLPLVSVSLFTPLAAEAPYMKRLSRQOVE 900
QY 901 DLEVLSDIDEMSRRRPELISFSTNLQRLMSAECCCNLAFFSLARSMONSPSTAAAF 960
Db 901 DLEVLSDIDEMSRRRPELISFSTNLQRLMSAECCCNLAFFSLARSMONSPSTAAAF 960
QY 961 LPTFMYCISQDPEVYOTALRNLPYALLCOEHAVALLHRAFLVGYGOMDPSAQISEAL 1020
Db 961 LPTFMYCISQDPEVYOTALRNLPYALLCOEHAVALLHRAFLVGYGOMDPSAQISEAL 1020
QY 1021 RILHMEAVM 1029
Db 1021 RILHMEAVM 1029

RESULT 5
US-10-033-435-22
: Sequence 22, Application US/10033435
: Publication No. US20030027256A1
: GENERAL INFORMATION:
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Goddard, Audrey
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Roy, Margaret Ann

```

```

APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Matanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2930R1CS
CURRENT APPLICATION NUMBER: US/10/033,435
CURRENT FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: 60/095,325
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/112,851
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,145
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,511
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/115,558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119,341
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119,537
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/119,965
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: 60/162,506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170,262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/187,202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/14941
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
SEQUENCE ID NO 22
LENGTH: 1029
TYPE: PRT
ORGANISM: Homo sapiens
US-10-033-435-22
Query Match 100.0%; Score 1029; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

121 DKNYMAHLYEYQHEBRCASGCGTFHSLTASLPDRDSTEAPKPSPEOPIGGRIRVGT 180
121 DKNYMAHLYEYQHEBRCASGCGTFHSLTASLPDRDSTEAPKPSPEOPIGGRIRVGT 180
181 QLRVLPEDDLAGMFLQIFPLSPDRMSSSPRYVALALQALQGLARVVGSPVEYGI 240
181 QLRVLPEDDLAGMFLQIFPLSPDRMSSSPRYVALALQALQGLARVVGSPVEYGI 240
241 TVRVQLATLTLSSPHGALVMSHRSHPFLACPLRLQCOQRKVPDDTGFSSFLVLL 300
241 TVRVQLATLTLSSPHGALVMSHRSHPFLACPLRLQCOQRKVPDDTGFSSFLVLL 300
301 QMLQMLDPSGVGGGRLRAQRLMLASQASAGRLSDVAGLRLALAFRODLEVSSTV 360
301 QMLQMLDPSGVGGGRLRAQRLMLASQASAGRLSDVAGLRLALAFRODLEVSSTV 360
361 RAVIATLRSGEGCSVEEDLISKVLQGLIEVRSPLBELTLAFSSATADAAFPACKPVV 420
361 RAVIATLRSGEGCSVEEDLISKVLQGLIEVRSPLBELTLAFSSATADAAFPACKPVV 420
421 VVSSLLLOEEEPFLAGKPGADGSLAVRUGPSSGLVDMLEMDPEVSSCPDLORL 480
421 VVSSLLLOEEEPFLAGKPGADGSLAVRUGPSSGLVDMLEMDPEVSSCPDLORL 480
481 FSRKRGQAQVPSFRPYLLTLFTHQSSWPTLHQCIRVLLGKSREORFDPASLDPLMAC 540
481 FSRKRGQAQVPSFRPYLLTLFTHQSSWPTLHQCIRVLLGKSREORFDPASLDPLMAC 540
541 IHVRIMQGRQRPQKRRRELVYRVGPELISVELLEAEFRSDGDTAAASLQAR 600
541 IHVRIMQGRQRPQKRRRELVYRVGPELISVELLEAEFRSDGDTAAASLQAR 600
601 LPILLSCCCGDESVRYVTEHLSGCIQWGDVYGRRCRDLLQLYQREPLRVPEVL 660
601 LPILLSCCCGDESVRYVTEHLSGCIQWGDVYGRRCRDLLQLYQREPLRVPEVL 660
661 LHSEGAASSVCKLDGLIHREITLADTSDSRALENRGADASMACRLAAVHPLLRHL 720
661 LHSEGAASSVCKLDGLIHREITLADTSDSRALENRGADASMACRLAAVHPLLRHL 720
721 PMIALHGRTHLNFQEFROONHLSCLHYVGLLELQHYFRSEHOGALMDCLSTIRL 780
721 PMIALHGRTHLNFQEFROONHLSCLHYVGLLELQHYFRSEHOGALMDCLSTIRL 780
781 LLYNRKSSRHIAFINKEVQFIHKYITYNAPAAISFLQKHADPLHDSFNSDLVMLKSL 840
781 LLYNRKSSRHIAFINKEVQFIHKYITYNAPAAISFLQKHADPLHDSFNSDLVMLKSL 840
841 LAGLSLPSRDDRDRLGDEBGEESAGSLPLVSVSLFTPLTAEMAAPYKRLSRGQTV 900
841 LAGLSLPSRDDRDRLGDEBGEESAGSLPLVSVSLFTPLTAEMAAPYKRLSRGQTV 900
901 DLEVLSDIDEMSRRELLISFSTNLQRLMSSAECCRNLAFLARSMONSPSTIAAF 960
901 DLEVLSDIDEMSRRELLISFSTNLQRLMSSAECCRNLAFLARSMONSPSTIAAF 960
961 LPTFMVCLGSODEEVQYOTALRNLPYVALLCOEHAVALHRAFLVGMYGQMDPSAQISEAL 1020
961 LPTFMVCLGSODEEVQYOTALRNLPYVALLCOEHAVALHRAFLVGMYGQMDPSAQISEAL 1020
1021 RIHMEAVM 1029
1021 RIHMEAVM 1029

```

```

RESULT 6
US-10-032-990-22
Sequence 22, Application US/10032990
Publication No. US20030032060A1
GENERAL INFORMATION:
APPLICANT: Botstein, David
APPLICANT: Desnovers, Luc

```

APPLICANT: Ferrara, Napoleone  
 APPLICANT: Fong, Sherman  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Pan, James  
 APPLICANT: Roy, Margaret Ann  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tamas, Daniel  
 APPLICANT: Watanabe, Colin K.  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 FILE REFERENCE: P2930R1C11  
 CURRENT FILING DATE: 2001-12-27  
 PRIOR APPLICATION NUMBER: 60/095,325  
 PRIOR FILING DATE: 1998-08-04  
 PRIOR APPLICATION NUMBER: 60/112,851  
 PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,145  
 PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,511  
 PRIOR FILING DATE: 1998-12-22  
 PRIOR APPLICATION NUMBER: 60/115,558  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,565  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,733  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/119,341  
 PRIOR FILING DATE: 1999-02-09  
 PRIOR APPLICATION NUMBER: 60/119,537  
 PRIOR FILING DATE: 1999-02-10  
 PRIOR APPLICATION NUMBER: 60/119,965  
 PRIOR FILING DATE: 1999-02-12  
 PRIOR APPLICATION NUMBER: 60/162,506  
 PRIOR FILING DATE: 1999-10-29  
 PRIOR APPLICATION NUMBER: 60/170,262  
 PRIOR FILING DATE: 1999-12-09  
 PRIOR APPLICATION NUMBER: 60/187,202  
 PRIOR FILING DATE: 2000-03-03  
 PRIOR APPLICATION NUMBER: PCT/US99/12252  
 PRIOR FILING DATE: 1999-06-02  
 PRIOR APPLICATION NUMBER: PCT/US99/28634  
 PRIOR FILING DATE: 1999-12-01  
 PRIOR APPLICATION NUMBER: PCT/US99/28551  
 PRIOR FILING DATE: 1999-12-02  
 PRIOR APPLICATION NUMBER: PCT/US00/03565  
 PRIOR FILING DATE: 2000-02-11  
 PRIOR APPLICATION NUMBER: PCT/US00/04414  
 PRIOR FILING DATE: 2000-02-22  
 PRIOR APPLICATION NUMBER: PCT/US00/05841  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: PCT/US00/08439  
 PRIOR FILING DATE: 2000-03-30  
 PRIOR APPLICATION NUMBER: PCT/US00/14941  
 PRIOR FILING DATE: 2000-05-30  
 PRIOR APPLICATION NUMBER: PCT/US00/15264  
 PRIOR FILING DATE: 2000-06-02  
 PRIOR APPLICATION NUMBER: PCT/US00/32678  
 PRIOR FILING DATE: 2000-12-01  
 NUMBER OF SEQ ID NOS: 38  
 SEQ ID NO 22  
 LENGTH: 1029  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-032-990-22

Query Match 100.0%; Score 1029; DB 9; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MHILVHAMVILITLGPRAADSEFOALDIMEPEEKPLPTAFIVDTSEALLPDMKL	60
DB	1	MHILVHAMVILITLGPRAADSEFOALDIMEPEEKPLPTAFIVDTSEALLPDMKL	60
QY	61	RMIRSEVLRVDAIADLEPOOLLFVOSFGIPVSSMSKLPLODAVAHQPTLEQNTM	120
DB	61	RMIRSEVLRVDAIADLEPOOLLFVOSFGIPVSSMSKLPLODAVAHQPTLEQNTM	120
QY	121	DKNYMAHLYEVOHERGASGGOTFHSILTASLPBRPSTAPPKSSPEPIGGRTIRVGT	180
DB	121	DKNYMAHLYEVOHERGASGGOTFHSILTASLPBRPSTAPPKSSPEPIGGRTIRVGT	180
QY	181	QLRVLGPEDDLAGMFLQIPIELSPDPWQSSPPVALALQALGQELAVVQSSPEVPGI	240
DB	181	QLRVLGPEDDLAGMFLQIPIELSPDPWQSSPPVALALQALGQELAVVQSSPEVPGI	240
QY	241	TVRVIALATLLSSPHGALVMSMRSHTLACPRLQOLCOYQRCVQDTGFSSFLTKVLL	300
DB	241	TVRVIALATLLSSPHGALVMSMRSHTLACPRLQOLCOYQRCVQDTGFSSFLTKVLL	300
QY	301	QMLQMDSPGVGGPLRAQLRMLASQASAGRISDVRGGLRLAEALARQDLEVVSTV	360
DB	301	QMLQMDSPGVGGPLRAQLRMLASQASAGRISDVRGGLRLAEALARQDLEVVSTV	360
QY	361	RAVIATLRSGECSVEPDLISKVQLIEVRSFHEBELLTAFPSATADAASPFPACKPVY	420
DB	361	RAVIATLRSGECSVEPDLISKVQLIEVRSFHEBELLTAFPSATADAASPFPACKPVY	420
QY	421	VVSSLLQEEEPRLAGKPGADGSLAVRLGPSGGLVWLEMLDPEVYSSCPDOLRL	480
DB	421	VVSSLLQEEEPRLAGKPGADGSLAVRLGPSGGLVWLEMLDPEVYSSCPDOLRL	480
QY	481	FSRRKKGQAQVPSFRPYLLTFTHOSSMPTLHQCIRVLLGSRREPRPSALDLMAC	540
DB	481	FSRRKKGQAQVPSFRPYLLTFTHOSSMPTLHQCIRVLLGSRREPRPSALDLMAC	540
QY	541	IVPRIMQGRDQTPQKRREELVLRVQPELISVLELLAEATRSQDDDTAACSILQAR	600
DB	541	IVPRIMQGRDQTPQKRREELVLRVQPELISVLELLAEATRSQDDDTAACSILQAR	600
QY	601	LPILLSCCGGDESVAKVTEHLSGCIQWGDVYLGRCDLLQLYLQREELRVPEYL	660
DB	601	LPILLSCCGGDESVAKVTEHLSGCIQWGDVYLGRCDLLQLYLQREELRVPEYL	660
QY	661	LHSEGAASSVCKLQGLIHRFTLLADTSDSRALENRGADASACRKLVAHPDLLRLH	720
DB	661	LHSEGAASSVCKLQGLIHRFTLLADTSDSRALENRGADASACRKLVAHPDLLRLH	720
QY	721	PMIAALLHGRTHLNFQEFROQNLSCFLVGLLELLQPHVFERSEHOGALMDCILSFIRL	780
DB	721	PMIAALLHGRTHLNFQEFROQNLSCFLVGLLELLQPHVFERSEHOGALMDCILSFIRL	780
QY	781	LNLYKRSRHLAFLINKVQFTHKYTYNAPAAISFLQKHADPLHDLSFSDNSDLVMLKSL	840
DB	781	LNLYKRSRHLAFLINKVQFTHKYTYNAPAAISFLQKHADPLHDLSFSDNSDLVMLKSL	840
QY	841	LAGLSIPSRDDRTDRGLDEGESEESAGSLPVSVSLFTPLAAEAPVPMKRISRCQTYE	900
DB	841	LAGLSIPSRDDRTDRGLDEGESEESAGSLPVSVSLFTPLAAEAPVPMKRISRCQTYE	900
QY	901	DLLEVLSDDIDEMSRREPELISFEFTNLQRLMSAECCCNLAFLSLARMQNSPSTAAAF	960
DB	901	DLLEVLSDDIDEMSRREPELISFEFTNLQRLMSAECCCNLAFLSLARMQNSPSTAAAF	960
QY	961	LPTFMVCLGSDPEFVQVTLRLNLPAYALICQEHAAVLLHRAFLVGMYGMDPSAQISEAL	1020
DB	961	LPTFMVCLGSDPEFVQVTLRLNLPAYALICQEHAAVLLHRAFLVGMYGMDPSAQISEAL	1020
QY	1021	RIIHEAVM 1029	
DB	1021	RIIHEAVM 1029	

```

RESULT 7
US-09-929-769-7
: Sequence 7, Application US/09929769
: Publication No: US20030055224A1
: GENERAL INFORMATION:
: APPLICANT: Gao, Mel-Olang
: APPLICANT: Polakis, Paul
: APPLICANT: Shou, Jianyong
: APPLICANT: Smith, Victoria
: APPLICANT: Soriano, Robert
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wu, Thomas D.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
: TITL OF INVENTION: TREATMENT OF TUMOR
: FILE REFERENCE: P5007R1-US
: CURRENT APPLICATION NUMBER: US/09/929, 769
: CURRENT FILING DATE: 2001-08-14
: PRIOR APPLICATION NUMBER: 60/089, 653
: PRIOR FILING DATE: 1998-06-17
: PRIOR APPLICATION NUMBER: 60/090, 355
: PRIOR FILING DATE: 1998-06-23
: PRIOR APPLICATION NUMBER: 60/104, 257
: PRIOR FILING DATE: 1998-10-14
: PRIOR APPLICATION NUMBER: 60/119, 537
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: 60/141, 037
: PRIOR FILING DATE: 1999-06-23
: PRIOR APPLICATION NUMBER: 60/145, 698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: 60/162, 506
: PRIOR FILING DATE: 1999-10-29
: PRIOR APPLICATION NUMBER: PCT/US99/12252
: PRIOR FILING DATE: 1999-06-02
: PRIOR APPLICATION NUMBER: PCT/US99/20111
: PRIOR FILING DATE: 1999-09-01
: PRIOR APPLICATION NUMBER: PCT/US99/28634
: PRIOR FILING DATE: 1999-12-01
: PRIOR APPLICATION NUMBER: PCT/US99/28551
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: PRIOR APPLICATION NUMBER: PCT/US00/00376
: PRIOR FILING DATE: 2000-01-06
: PRIOR APPLICATION NUMBER: PCT/US00/04342
: PRIOR FILING DATE: 2000-02-18
: PRIOR APPLICATION NUMBER: PCT/US00/08439
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: PCT/US00/13705
: PRIOR FILING DATE: 2000-05-17
: PRIOR APPLICATION NUMBER: PCT/US00/23328
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: PCT/US01/06520
: PRIOR FILING DATE: 2001-02-28
: PRIOR APPLICATION NUMBER: PCT/US01/20118
: PRIOR FILING DATE: 2001-06-22
: PRIOR APPLICATION NUMBER: 09/888, 257
: PRIOR FILING DATE: 2001-06-22
: NUMBER OF SEQ ID NOS: 8
: SEQ ID NO 7
: LENGTH: 1029
: TYPE: PRT
: ORGANISM: Homo Sapien
US-09-929-769-7

Query Match      100.0%; Score 1029; DB 9; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 MHLVVHNAVYILTLGPRRADSEFOALDIWPEEKPRTAFVDTSEALILPWLKL 60
DB 1 MHLVVHNAVYILTLGPRRADSEFOALDIWPEEKPRTAFVDTSEALILPWLKL 60

```

```

OY 61 RMIRSEVLRLVDALQDLEPOQLLEFVQSGFIPVSSSKLLQFLDQAVHPDQTLQONIM 120
DB 61 RMIRSEVLRLVDALQDLEPOQLLEFVQSGFIPVSSSKLLQFLDQAVHPDQTLQONIM 120
OY 121 DKNYMAHLVEQVHERGASGQTFHSLLTASLPPRRDSTEAPKPKSSPEOPIGGRIINVT 180
DB 121 DKNYMAHLVEQVHERGASGQTFHSLLTASLPPRRDSTEAPKPKSSPEOPIGGRIINVT 180
OY 181 QLRVLGPEDDLACMFQILFPLSPDRMOWSSPRVALALQALQOELARVQGSPEVPGI 240
DB 181 QLRVLGPEDDLACMFQILFPLSPDRMOWSSPRVALALQALQOELARVQGSPEVPGI 240
OY 241 TVRYLQALATLLSSPHGALVMSHRSHPFLACPLLRQLCOYRCVPDGTGFSFLTKVLL 300
DB 241 TVRYLQALATLLSSPHGALVMSHRSHPFLACPLLRQLCOYRCVPDGTGFSFLTKVLL 300
OY 301 OMLOWMDSPGVEGGPLRAOLRLMASQASAGRISDVGGILRLAEALAFQDLEVSSTV 360
DB 301 OMLOWMDSPGVEGGPLRAOLRLMASQASAGRISDVGGILRLAEALAFQDLEVSSTV 360
OY 361 RAVIATLRSGEQCSVEPDLISKVQLIEVRSPLLELTAFTSATDAASPPACKPVY 420
DB 361 RAVIATLRSGEQCSVEPDLISKVQLIEVRSPLLELTAFTSATDAASPPACKPVY 420
OY 421 VVSSLLLQEEEPFLAGKRGADGSGLEAVRIGPSSGGLVMDLEMDPEVSSCPDLQRL 480
DB 421 VVSSLLLQEEEPFLAGKRGADGSGLEAVRIGPSSGGLVMDLEMDPEVSSCPDLQRL 480
OY 481 FSRKRGQAQVPSFRPYLLTLFTHOSWPTLHQCIRVLGKSGREQFDSASLDFLMAC 540
DB 481 FSRKRGQAQVPSFRPYLLTLFTHOSWPTLHQCIRVLGKSGREQFDSASLDFLMAC 540
OY 541 IHVPRIWQGRDQTPQKRREBELVRYOGPELISVELITAEATRSODSGTAACSLIQAR 600
DB 541 IHVPRIWQGRDQTPQKRREBELVRYOGPELISVELITAEATRSODSGTAACSLIQAR 600
OY 601 LPLLLSCCGDDSVKRVTEHLISGCIQOWKDSVYGRRCRLQLQLYORELRYPEVEVL 660
DB 601 LPLLLSCCGDDSVKRVTEHLISGCIQOWKDSVYGRRCRLQLQLYORELRYPEVEVL 660
OY 661 LHSEGAASSVCKLDGLIHRFTLLADTSDSRALENRGADASMACRLAVANHLPLLRHL 720
DB 661 LHSEGAASSVCKLDGLIHRFTLLADTSDSRALENRGADASMACRLAVANHLPLLRHL 720
OY 721 PMTALILHGRTHLNFQFROONHLSCLHYLGLLELLOPHVFSEHOGALMDCLSTR 780
DB 721 PMTALILHGRTHLNFQFROONHLSCLHYLGLLELLOPHVFSEHOGALMDCLSTR 780
OY 781 LNAVYKSSRLAAFINFEVQFIHKYIYNNAPAAISFLQKHADPLHDSFNSDLYMKSL 840
DB 781 LNAVYKSSRLAAFINFEVQFIHKYIYNNAPAAISFLQKHADPLHDSFNSDLYMKSL 840
OY 841 LAGLSLPSRDDRTDRCGLDEGESESSAGSLPLVSVSLFTPLTAEMA PYMKRLSRGQTV 900
DB 841 LAGLSLPSRDDRTDRCGLDEGESESSAGSLPLVSVSLFTPLTAEMA PYMKRLSRGQTV 900
OY 901 DLLEVLSDIDEMSRREILISFSTNLQRLMSSAEECCRNLAFLARSQNSPSTAAAF 960
DB 901 DLLEVLSDIDEMSRREILISFSTNLQRLMSSAEECCRNLAFLARSQNSPSTAAAF 960
OY 961 LPTFMVCLGSDQFEVYOTALRNLPEYALLQCEHAAVYLHRAFLVGMGQMDPSAQISEAL 1020
DB 961 LPTFMVCLGSDQFEVYOTALRNLPEYALLQCEHAAVYLHRAFLVGMGQMDPSAQISEAL 1020
OY 1021 RILHMEAVM 1029
DB 1021 RILHMEAVM 1029

```

```

RESULT 8
US-10-032-996-22
: Sequence 22, Application US/10032996
: Publication No. US20030054447A1

```

GENERAL INFORMATION:  
 APPLICANT: Botstein, David  
 APPLICANT: Desnoyers, Luc  
 APPLICANT: Ferrara, Napoleone  
 APPLICANT: Fong, Sherman  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Pan, James  
 APPLICANT: Roy, Margaret Ann  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumas, Daniel  
 APPLICANT: Watanabe, Colin K.  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 TITLE OF INVENTION: Acids Encoding the Same  
 FILE REFERENCE: P2930R1C3  
 CURRENT APPLICATION NUMBER: US/10/032,996  
 CURRENT FILING DATE: 2001-12-27  
 PRIOR APPLICATION NUMBER: 60/095,325  
 PRIOR FILING DATE: 1998-08-04  
 PRIOR APPLICATION NUMBER: 60/112,851  
 PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,145  
 PRIOR FILING DATE: 1998-12-16  
 PRIOR APPLICATION NUMBER: 60/113,511  
 PRIOR FILING DATE: 1998-12-22  
 PRIOR APPLICATION NUMBER: 60/115,558  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,565  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115,733  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/119,341  
 PRIOR FILING DATE: 1999-02-09  
 PRIOR APPLICATION NUMBER: 60/119,537  
 PRIOR FILING DATE: 1999-02-10  
 PRIOR APPLICATION NUMBER: 60/119,965  
 PRIOR FILING DATE: 1999-02-12  
 PRIOR APPLICATION NUMBER: 60/162,506  
 PRIOR FILING DATE: 1999-10-29  
 PRIOR APPLICATION NUMBER: 60/170,262  
 PRIOR FILING DATE: 1999-12-09  
 PRIOR APPLICATION NUMBER: 60/187,202  
 PRIOR FILING DATE: 2000-03-03  
 PRIOR APPLICATION NUMBER: PCT/US99/12252  
 PRIOR FILING DATE: 1999-06-02  
 PRIOR APPLICATION NUMBER: PCT/US99/28634  
 PRIOR FILING DATE: 1999-12-01  
 PRIOR APPLICATION NUMBER: PCT/US99/28551  
 PRIOR FILING DATE: 1999-12-02  
 PRIOR APPLICATION NUMBER: PCT/US00/03565  
 PRIOR FILING DATE: 2000-02-11  
 PRIOR APPLICATION NUMBER: PCT/US00/04414  
 PRIOR FILING DATE: 2000-02-22  
 PRIOR APPLICATION NUMBER: PCT/US00/05841  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: PCT/US00/08439  
 PRIOR FILING DATE: 2000-03-30  
 PRIOR APPLICATION NUMBER: PCT/US00/14941  
 PRIOR FILING DATE: 2000-05-30  
 PRIOR APPLICATION NUMBER: PCT/US00/15264  
 PRIOR FILING DATE: 2000-06-02  
 PRIOR APPLICATION NUMBER: PCT/US00/32678  
 PRIOR FILING DATE: 2000-12-01  
 NUMBER OF SEQ ID NOS: 38  
 SEQ ID NO 22  
 LENGTH: 1029  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-032-996-22

Query Match 100.0% Score 1029, DB 9, Length 1029;

Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MHLYVHAHVILTLTPRRADSEFOALDIWPEKPLPTALVLTSEALLPWL 60  
 DB 1 MHLYVHAHVILTLTPRRADSEFOALDIWPEKPLPTALVLTSEALLPWL 60  
 QY 61 RMIRSEVLRVDAALDLEPQOLLFPVSGIVSSMSKILQPLDVAADPOTLCON 120  
 DB 61 RMIRSEVLRVDAALDLEPQOLLFPVSGIVSSMSKILQPLDVAADPOTLCON 120  
 QY 121 DKNYMAHVEVOHERGASGGQTFHSLTASLPRRSTEARPKRSSPEQIGGRIRVGT 180  
 DB 121 DKNYMAHVEVOHERGASGGQTFHSLTASLPRRSTEARPKRSSPEQIGGRIRVGT 180  
 QY 181 QLRVLGPEDDLAGMFLQIFPLSPDPWQSSPPVALALQALGQELARVQCSPEVPGI 240  
 DB 181 QLRVLGPEDDLAGMFLQIFPLSPDPWQSSPPVALALQALGQELARVQCSPEVPGI 240  
 QY 241 TVRVLOALATLSSPHGALVMSHRSHPFLACPLRQLQYORCVPODTGFSFLKVL 300  
 DB 241 TVRVLOALATLSSPHGALVMSHRSHPFLACPLRQLQYORCVPODTGFSFLKVL 300  
 QY 301 OMLQWLDSPEVEGGLRAQLRMLASQASAGRRLSDVRCGLLRRLAEALAEARQDELEVSTV 360  
 DB 301 OMLQWLDSPEVEGGLRAQLRMLASQASAGRRLSDVRCGLLRRLAEALAEARQDELEVSTV 360  
 QY 361 RAVIATLRSGECCVPEPDLISKVLOGLIEYRSHLEBELTAFTSATAADAAPPACKPVY 420  
 DB 361 RAVIATLRSGECCVPEPDLISKVLOGLIEYRSHLEBELTAFTSATAADAAPPACKPVY 420  
 QY 421 VVSSLLQEEEPPLAGKPGADGSLFAVRIGPSSGLVMDLEMDPEVSSCDLQRL 480  
 DB 421 VVSSLLQEEEPPLAGKPGADGSLFAVRIGPSSGLVMDLEMDPEVSSCDLQRL 480  
 QY 481 FSRKRGKGAQVPSFRPYLTLLFTTHQSWPTLQCIKVLGKREGRFSPASLDELWAC 540  
 DB 481 FSRKRGKGAQVPSFRPYLTLLFTTHQSWPTLQCIKVLGKREGRFSPASLDELWAC 540  
 QY 541 IHVPRIMOGRDQTPQKRREELVLRVQGPPELISLVELILAEATRSDGDTAACSILQAR 600  
 DB 541 IHVPRIMOGRDQTPQKRREELVLRVQGPPELISLVELILAEATRSDGDTAACSILQAR 600  
 QY 601 LPLLLSCCGGDDSVKRVTEHLSGCIQOMGDSVLRGRCDLQLVLRQRELVPVPEYL 660  
 DB 601 LPLLLSCCGGDDSVKRVTEHLSGCIQOMGDSVLRGRCDLQLVLRQRELVPVPEYL 660  
 QY 661 LHSEGAASSVCKLDGLIHRFTLLADTSDRALENRGADASNAKRLAVAHPLLLRL 720  
 DB 661 LHSEGAASSVCKLDGLIHRFTLLADTSDRALENRGADASNAKRLAVAHPLLLRL 720  
 QY 721 PMTALLHGRTHNPODFRQONHLSGFHYVLGELLELOPHVFRSEHOGALMDCLLSFIRL 780  
 DB 721 PMTALLHGRTHNPODFRQONHLSGFHYVLGELLELOPHVFRSEHOGALMDCLLSFIRL 780  
 QY 781 LNTYKSSRHIAAFINKFVOFIRKITYTYPAPAISELOKHADPLHLSFONSJLVMLKSL 840  
 DB 781 LNTYKSSRHIAAFINKFVOFIRKITYTYPAPAISELOKHADPLHLSFONSJLVMLKSL 840  
 QY 841 LAGLSLPSRDDRTRGLDEBGEESAGSLPLVSVSLTPTLAEMAPYKRLRSQGYE 900  
 DB 841 LAGLSLPSRDDRTRGLDEBGEESAGSLPLVSVSLTPTLAEMAPYKRLRSQGYE 900  
 QY 901 DLLEVLSDIDEMRRRPELISFSTNLQRLMSAECCRLASLRLSOMNSPSTAAAF 960  
 DB 901 DLLEVLSDIDEMRRRPELISFSTNLQRLMSAECCRLASLRLSOMNSPSTAAAF 960  
 QY 961 LPTFMVCLSGODEVEVQTLARNLPETALLCOEHAVALIHRAFIVGYMGQDPDAQISEAL 1020  
 DB 961 LPTFMVCLSGODEVEVQTLARNLPETALLCOEHAVALIHRAFIVGYMGQDPDAQISEAL 1020  
 QY 1021 RILHMEAVM 1029  
 DB 1021 RILHMEAVM 1029



Db 1021 RILHMEAVM 1029

## RESULT 9

US-10-033-396-22  
 ; Sequence 22, Application US/10033396  
 ; Publication No. US20030077657A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Botstein, David  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Ferrara, Napoleone  
 ; APPLICANT: Fong, Sherman  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Roy, Margaret Ann  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Collin K.  
 ; APPLICANT: Wood, William I.  
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 ; FILE REFERENCE: P2930R1C4  
 ; CURRENT APPLICATION NUMBER: US/10/033,396  
 ; CURRENT FILING DATE: 2001-12-27  
 ; PRIOR APPLICATION NUMBER: 60/095,325  
 ; PRIOR FILING DATE: 1998-08-04  
 ; PRIOR APPLICATION NUMBER: 60/112,851  
 ; PRIOR FILING DATE: 1998-12-16  
 ; PRIOR APPLICATION NUMBER: 60/113,145  
 ; PRIOR FILING DATE: 1998-12-16  
 ; PRIOR APPLICATION NUMBER: 60/113,511  
 ; PRIOR FILING DATE: 1998-12-22  
 ; PRIOR APPLICATION NUMBER: 60/115,558  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/115,565  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/115,733  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/119,341  
 ; PRIOR FILING DATE: 1999-02-09  
 ; PRIOR APPLICATION NUMBER: 60/119,537  
 ; PRIOR FILING DATE: 1999-02-10  
 ; PRIOR APPLICATION NUMBER: 60/119,965  
 ; PRIOR FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: 60/162,506  
 ; PRIOR FILING DATE: 1999-10-29  
 ; PRIOR APPLICATION NUMBER: 60/170,262  
 ; PRIOR FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: 60/187,202  
 ; PRIOR FILING DATE: 2000-03-03  
 ; PRIOR APPLICATION NUMBER: PCT/US99/12252  
 ; PRIOR FILING DATE: 1999-06-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28634  
 ; PRIOR FILING DATE: 1999-12-01  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28551  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US00/03565  
 ; PRIOR FILING DATE: 2000-02-11  
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414  
 ; PRIOR FILING DATE: 2000-02-22  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05841  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: PCT/US00/08439  
 ; PRIOR FILING DATE: 2000-03-30  
 ; PRIOR APPLICATION NUMBER: PCT/US00/14941  
 ; PRIOR FILING DATE: 2000-05-30  
 ; PRIOR APPLICATION NUMBER: PCT/US00/15264  
 ; PRIOR FILING DATE: 2000-06-02  
 ; PRIOR APPLICATION NUMBER: PCT/US00/32678  
 ; PRIOR FILING DATE: 2000-12-01  
 ; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 22  
 ; LENGTH: 1029  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-033-396-22  
 Query Match 100.0%; Score 1029; DB 9; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MHILVHAMVILLTLGPRADSEFOALDITMPEKRLPTAFIVDSEELLLPDWIKL	60
Db	1	MHILVHAMVILLTLGPRADSEFOALDITMPEKRLPTAFIVDSEELLLPDWIKL	60
Qy	61	RMIRSEVRLVDALDLEPOOLLLFVOSFGIPVSSMSKLLQFLDQAAVHPDTEQNIM	120
Db	61	RMIRSEVRLVDALDLEPOOLLLFVOSFGIPVSSMSKLLQFLDQAAVHPDTEQNIM	120
Qy	121	DKNTMAHLEVQHERGASGGTFFHSLTASLPDRDSTEAPKPKSPQPIGQGRIRVGT	180
Db	121	DKNTMAHLEVQHERGASGGTFFHSLTASLPDRDSTEAPKPKSPQPIGQGRIRVGT	180
Qy	181	QLRVIGPEDDLAGMFLQIFPLSPDRMOSSPRPVALALQOALGQELARVVGSEVPGI	240
Db	181	QLRVIGPEDDLAGMFLQIFPLSPDRMOSSPRPVALALQOALGQELARVVGSEVPGI	240
Qy	241	TVRVLOALATLLSSPHGALVMSHRSHTLACPLLRQLCOYQRCVPQDTGFSSTFLKVL	300
Db	241	TVRVLOALATLLSSPHGALVMSHRSHTLACPLLRQLCOYQRCVPQDTGFSSTFLKVL	300
Qy	301	QMLQWLDSPGVGGPLRAQLRLASQASAGRLLSVRGGLRLAALAFRODLEVVSSTV	360
Db	301	QMLQWLDSPGVGGPLRAQLRLASQASAGRLLSVRGGLRLAALAFRODLEVVSSTV	360
Qy	361	RAVATLRSQGCSVEPRLISKVLQGLTEVRSPHLEELLTFEFSATDAASPFPACKPVV	420
Db	361	RAVATLRSQGCSVEPRLISKVLQGLTEVRSPHLEELLTFEFSATDAASPFPACKPVV	420
Qy	421	VVSSLLQEEBPLAGKRGADGSLAVALRGPSSGLVDWMLDPEVSSCPDQLRL	480
Db	421	VVSSLLQEEBPLAGKRGADGSLAVALRGPSSGLVDWMLDPEVSSCPDQLRL	480
Qy	481	FSRRKGKQAOVPSFRPLTLFTHQSSWPLTHQCIIRVLGKSRQRDPASLDFLWAC	540
Db	481	FSRRKGKQAOVPSFRPLTLFTHQSSWPLTHQCIIRVLGKSRQRDPASLDFLWAC	540
Qy	541	IHPRIHOGROOFTPOKREELVLRVGPPELISVELLIAEFTSQGDPAACSLIOAR	600
Db	541	IHPRIHOGROOFTPOKREELVLRVGPPELISVELLIAEFTSQGDPAACSLIOAR	600
Qy	601	LPLLSCCGDDESVRKVTEHLSCGIQMGDSVIGRRCDLLDLQRPFLRVVPEVL	660
Db	601	LPLLSCCGDDESVRKVTEHLSCGIQMGDSVIGRRCDLLDLQRPFLRVVPEVL	660
Qy	661	LHSEGAASSVCKLDGLIHRFTITLADTSDSRALENKRAADSMACRKLAVAHPILLHL	720
Db	661	LHSEGAASSVCKLDGLIHRFTITLADTSDSRALENKRAADSMACRKLAVAHPILLHL	720
Qy	721	PMIAALLGFTHLNFOEPRQONHLSCLFHVGLLELLOPHYFRSHQALWDCLISFTRL	780
Db	721	PMIAALLGFTHLNFOEPRQONHLSCLFHVGLLELLOPHYFRSHQALWDCLISFTRL	780
Qy	781	LLNTRKSSRHIAAFINKVQETIHKYTYTNAPAAISFLQKHADPLHDSFDNSDLVLMKSL	840
Db	781	LLNTRKSSRHIAAFINKVQETIHKYTYTNAPAAISFLQKHADPLHDSFDNSDLVLMKSL	840
Qy	841	LAGLSLPSRDRTDGLDEGEESAGSLPLVYSVLTPLTAABMAYMKRLSGQTV	900
Db	841	LAGLSLPSRDRTDGLDEGEESAGSLPLVYSVLTPLTAABMAYMKRLSGQTV	900
Qy	901	DLLEVLSDIDMSRRRPETLSFFSTNLORLMSAECCRNLAFAISLALRSMNSPSIAAF	960
Db	901	DLLEVLSDIDMSRRRPETLSFFSTNLORLMSAECCRNLAFAISLALRSMNSPSIAAF	960



Qy 961 LPTFMVCLGSDPEFVYQVATLRNLPEYALICQEHAAVLHRAFLVGYGMQDPSAQISEAL 1020  
 |||  
 Db 961 LPTFMVCLGSDPEFVYQVATLRNLPEYALICQEHAAVLHRAFLVGYGMQDPSAQISEAL 1020  
 |||  
 Qy 1021 RLHMEAVM 1029  
 |||  
 Db 1021 RLHMEAVM 1029  
 |||  
 RESULT 10  
 US-10-033-246-22  
 ; Sequence 22, Application US/10033246  
 ; Patent No. US20020098505A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Botstein, David  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Ferrara, Napoleone  
 ; APPLICANT: Fong, Sherman  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Roy, Margaret Ann  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 ; FILE OF INVENTION: Acids Encoding the Same  
 ; FILE REFERENCE: P2930R1C12  
 ; CURRENT FILING DATE: 2001-12-28  
 ; PRIOR APPLICATION NUMBER: 60/095,325  
 ; PRIOR FILING DATE: 1998-08-04  
 ; PRIOR APPLICATION NUMBER: 60/112,851  
 ; PRIOR FILING DATE: 1998-12-16  
 ; PRIOR APPLICATION NUMBER: 60/113,145  
 ; PRIOR FILING DATE: 1998-12-16  
 ; PRIOR APPLICATION NUMBER: 60/113,511  
 ; PRIOR FILING DATE: 1998-12-22  
 ; PRIOR APPLICATION NUMBER: 60/115,558  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/115,565  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/115,733  
 ; PRIOR FILING DATE: 1999-01-12  
 ; PRIOR APPLICATION NUMBER: 60/119,341  
 ; PRIOR FILING DATE: 1999-02-09  
 ; PRIOR APPLICATION NUMBER: 60/119,537  
 ; PRIOR FILING DATE: 1999-02-10  
 ; PRIOR APPLICATION NUMBER: 60/119,965  
 ; PRIOR FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: 60/162,506  
 ; PRIOR FILING DATE: 1999-10-29  
 ; PRIOR APPLICATION NUMBER: 60/170,262  
 ; PRIOR FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: 60/187,202  
 ; PRIOR FILING DATE: 2000-03-03  
 ; PRIOR APPLICATION NUMBER: PCT/US99/12252  
 ; PRIOR FILING DATE: 1999-06-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28634  
 ; PRIOR FILING DATE: 1999-12-01  
 ; PRIOR APPLICATION NUMBER: PCT/US99/28551  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: PCT/US00/03565  
 ; PRIOR FILING DATE: 2000-02-11  
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414  
 ; PRIOR FILING DATE: 2000-02-22  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05841  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: PCT/US00/08439  
 ; PRIOR FILING DATE: 2000-03-30

;; PRIOR APPLICATION NUMBER: PCT/US00/14941  
 ; PRIOR FILING DATE: 2000-05-30  
 ; PRIOR APPLICATION NUMBER: PCT/US00/15264  
 ; PRIOR FILING DATE: 2000-06-02  
 ; PRIOR APPLICATION NUMBER: PCT/US00/32678  
 ; PRIOR FILING DATE: 2000-12-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 22  
 ; LENGTH: 1029  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-033-246-22  
 Query Match 100.0%; Score 1029; DB 12; Length 1029;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MHILVHAMVILTLGPPRADSEFOALLDIWPEEKPLPTAFVDTSEBALLPDMKL 60  
 |||  
 Db 1 MHILVHAMVILTLGPPRADSEFOALLDIWPEEKPLPTAFVDTSEBALLPDMKL 60  
 |||  
 Qy 61 RMIRSEVLRVDAALQDLPEQQLLEFVQSGFLPVSSMSKLLQFLDQAVAHDPQTLQONIM 120  
 |||  
 Db 61 RMIRSEVLRVDAALQDLPEQQLLEFVQSGFLPVSSMSKLLQFLDQAVAHDPQTLQONIM 120  
 |||  
 Qy 121 DKNYMAHVEVOHERGASGGQTFHSILITASLPERRDSTEAPKPKSSPEOPICGRIYGT 180  
 |||  
 Db 121 DKNYMAHVEVOHERGASGGQTFHSILITASLPERRDSTEAPKPKSSPEOPICGRIYGT 180  
 |||  
 Qy 181 QLRVLGPEDDLQAGMFLQIFPLSPDPMWSSPPRVALLQALGQELARVQSSPEVPGI 240  
 |||  
 Db 181 QLRVLGPEDDLQAGMFLQIFPLSPDPMWSSPPRVALLQALGQELARVQSSPEVPGI 240  
 |||  
 Qy 241 TVRVLQALATLSSPHGALVMSHRSHPACGLLQLOQYQYCVQDQCFSSFLKVL 300  
 |||  
 Db 241 TVRVLQALATLSSPHGALVMSHRSHPACGLLQLOQYQYCVQDQCFSSFLKVL 300  
 |||  
 Qy 301 QMLQWLDSPGVEGPIRLAQLRMLASQASAGRRLSDVRGGLRLRAELAFRODLEVVSTY 360  
 |||  
 Db 301 QMLQWLDSPGVEGPIRLAQLRMLASQASAGRRLSDVRGGLRLRAELAFRODLEVVSTY 360  
 |||  
 Qy 361 RAVIATLRSGEGCVPEPDLISKVLQGLIEVRSFHEELLTAFFSAPADAAPPACKPVY 420  
 |||  
 Db 361 RAVIATLRSGEGCVPEPDLISKVLQGLIEVRSFHEELLTAFFSAPADAAPPACKPVY 420  
 |||  
 Qy 421 VVSSILLQEEPLAGGKPGADGSLFAVRLGPPSSGLVWMLDPEVVSQCDLQRL 480  
 |||  
 Db 421 VVSSILLQEEPLAGGKPGADGSLFAVRLGPPSSGLVWMLDPEVVSQCDLQRL 480  
 |||  
 Qy 481 FSRKKGQAQVPSFRPYLLTLFTHQSMPTLHQCIRVLGKSGREORFPASIDELMAC 540  
 |||  
 Db 481 FSRKKGQAQVPSFRPYLLTLFTHQSMPTLHQCIRVLGKSGREORFPASIDELMAC 540  
 |||  
 Qy 541 IHVPRIWQGDORTPOKRRRELVLRVQGPFLISLVELLIAEATRSODGDTAACSILQAR 600  
 |||  
 Db 541 IHVPRIWQGDORTPOKRRRELVLRVQGPFLISLVELLIAEATRSODGDTAACSILQAR 600  
 |||  
 Qy 601 LPFLSCCGGDDSVKRVTEHLSGCIOQMGDSVLGRRCRDLLQTLQRELVVPEVYL 660  
 |||  
 Db 601 LPFLSCCGGDDSVKRVTEHLSGCIOQMGDSVLGRRCRDLLQTLQRELVVPEVYL 660  
 |||  
 Qy 661 LHSGEAASSVCKLDGLIHREFTLADTSDSRALENKGDASMACRKLAVAHPLLLRL 720  
 |||  
 Db 661 LHSGEAASSVCKLDGLIHREFTLADTSDSRALENKGDASMACRKLAVAHPLLLRL 720  
 |||  
 Qy 721 PMIAALLHGRTHLNFOEPRQONHLSFVHVLGLLELLQPRVFSSEHGALMDLSEFIRL 780  
 |||  
 Db 721 PMIAALLHGRTHLNFOEPRQONHLSFVHVLGLLELLQPRVFSSEHGALMDLSEFIRL 780  
 |||  
 Qy 781 LKNYKSSRHIAFINKFQFIHKYTTYNAIPAISFLQKHADPLHLSFSDNSDLVMLKSL 840  
 |||  
 Db 781 LKNYKSSRHIAFINKFQFIHKYTTYNAIPAISFLQKHADPLHLSFSDNSDLVMLKSL 840  
 |||

```

QY 841 LAGISLPSRDDRDGLDEBGESESSAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQTV 900
DB 841 LAGISLPSRDDRDRLGRGLDEBGESESSAGSLPLVSVSLFTPLTAEMAPYMKRLSRGQTV 900
QY 901 DLEVLSDIDEMSRREPELISFSTNLQRLMSSAECCORNLAFSLARSMQNSISIAAF 960
DB 901 DLEVLSDIDEMSRREPELISFSTNLQRLMSSAECCORNLAFSLARSMQNSISIAAF 960
QY 961 LPTFMVGLSGDSFVQVOTALRNLPEVALLCOEHAVALLHRAFLVGMYGOMDPSAQISEAL 1020
DB 961 LPTFMVGLSGDSFVQVOTALRNLPEVALLCOEHAVALLHRAFLVGMYGOMDPSAQISEAL 1020
QY 1021 RLHMEAVM 1029
DB 1021 RLHMEAVM 1029

RESULT 11
US-10-033-301-22
; Sequence 22, Application US/10033301
; Patent No. US20020098506A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2930R1C6
; CURRENT APPLICATION NUMBER: US/10/033,301
; CURRENT FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/095,325
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/112,851
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,145
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,511
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/115,558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119,341
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119,965
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/187,202
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/03565

```

```

; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 22
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-033-301-22

Query Match 100.0%; Score 1029; DB 12; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHILVHAMVILLTLGPPRADSEFOALLDIWPEEKPLPAFLVDTSSEALLLPDWLKL 60
DB 1 MHILVHAMVILLTLGPPRADSEFOALLDIWPEEKPLPAFLVDTSSEALLLPDWLKL 60
QY 61 RMIRSEVRLVDALQDLEPOQLLFVQSGFIPVSSMSKLLQFLDOVAHHPOTLEONIM 120
DB 61 RMIRSEVRLVDALQDLEPOQLLFVQSGFIPVSSMSKLLQFLDOVAHHPOTLEONIM 120
QY 121 DKNYMAHLYEVQHERGASGGOTFHSLLTASLPRRDSTEAPKPSPEOPIGGRIRYGT 180
DB 121 DKNYMAHLYEVQHERGASGGOTFHSLLTASLPRRDSTEAPKPSPEOPIGGRIRYGT 180
QY 121 DKNYMAHLYEVQHERGASGGOTFHSLLTASLPRRDSTEAPKPSPEOPIGGRIRYGT 180
DB 121 DKNYMAHLYEVQHERGASGGOTFHSLLTASLPRRDSTEAPKPSPEOPIGGRIRYGT 180
QY 181 QLRVLGPEDDLDAGMFLQFLFPLSPDRMSSSPRYALALQALQGLARVQSGSEVYGI 240
DB 181 QLRVLGPEDDLDAGMFLQFLFPLSPDRMSSSPRYALALQALQGLARVQSGSEVYGI 240
QY 241 TVRYLQALATLTLSSPHGALVMSMHRSHFLACPLLRLQCYQRCVPODTGFSFLKYL 300
DB 241 TVRYLQALATLTLSSPHGALVMSMHRSHFLACPLLRLQCYQRCVPODTGFSFLKYL 300
QY 301 QMLQWLSPEVGEGLRLRQLMLASQASAGRRSLDVREGGLRLALAFRODLEVYSTV 360
DB 301 QMLQWLSPEVGEGLRLRQLMLASQASAGRRSLDVREGGLRLALAFRODLEVYSTV 360
QY 361 RAVATATLRSGGOCVEPDLISKVLQGLLEVRSPLHEELLTAFFSATADAAAPFPACKRYV 420
DB 361 RAVATATLRSGGOCVEPDLISKVLQGLLEVRSPLHEELLTAFFSATADAAAPFPACKRYV 420
QY 421 VVSSLILQEEPEPLAGKKGADGSLAVRLGSSGLLVDMLEMDPEVSSCPDLQRL 480
DB 421 VVSSLILQEEPEPLAGKKGADGSLAVRLGSSGLLVDMLEMDPEVSSCPDLQRL 480
QY 481 FSRKRGKQAOVPSFRPYLLFLTHQSSWPTLHQCIRYLKSKRQRPDPASLDFLMAC 540
DB 481 FSRKRGKQAOVPSFRPYLLFLTHQSSWPTLHQCIRYLKSKRQRPDPASLDFLMAC 540
QY 541 IHVPRIMGROORTPOKRRRELVLRVQGPFLISVELILAAETRSQGDPAACSLIOAR 600
DB 541 IHVPRIMGROORTPOKRRRELVLRVQGPFLISVELILAAETRSQGDPAACSLIOAR 600
QY 601 LPLILSCCGGDESVRYKTEHLSGCIQOWGDSVLRRCRDLLDLQYLRPELRYVPEVL 660
DB 601 LPLILSCCGGDESVRYKTEHLSGCIQOWGDSVLRRCRDLLDLQYLRPELRYVPEVL 660
QY 661 LHSEGAASSVCKLDGLTHRFITLLADTSDSRALENRGADASMACRKLAVAHPLLLRL 720
DB 661 LHSEGAASSVCKLDGLTHRFITLLADTSDSRALENRGADASMACRKLAVAHPLLLRL 720
QY 721 PMIAALLHGRTHLNFQERQONHLSCFLHVLGILELLQPHVFRSRHQALMDCLLSFRL 780

```

```

Db 721 PMIAALLGRTHLNFQEEFQOHNLSCEFLVGLLELLQPHVFRSEHOGALMPCLLSIFRL 780
Qy 781 LLYAKSSSHLAAFINFKYQFTHKITYTNAAPAI5FLOKHAPRLDLSNDSDLVWKL 840
Db 781 LLYAKSSSHLAAFINFKYQFTHKITYTNAAPAI5FLOKHAPRLDLSNDSDLVWKL 840
Qy 841 LAGLSRDRDRTDGLDEEGEESAGSLPVSVSLFTPLAAEMAPMKLSRQOY 900
Db 841 LAGLSRDRDRTDGLDEEGEESAGSLPVSVSLFTPLAAEMAPMKLSRQOY 900
Qy 901 DLEVLSDIDEMSRRRPELISFFSTNLORLMSAECCRNIAFLALRSMONSPSIAAF 960
Db 901 DLEVLSDIDEMSRRRPELISFFSTNLORLMSAECCRNIAFLALRSMONSPSIAAF 960
Qy 961 LPTFNVCGSODEFVYVOTARMLPEYALICQEHAAVLLHRAFLVGYGMQDSQAQISEAL 1020
Db 961 LPTFNVCGSODEFVYVOTARMLPEYALICQEHAAVLLHRAFLVGYGMQDSQAQISEAL 1020
Qy 1021 RILHMEAVM 1029
Db 1021 RILHMEAVM 1029

```

```

RESULT 12
US-10-033-326-22
Sequence 22, Application US/10033326
Patent No. US20020098507A1
GENERAL INFORMATION:
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tamas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2930R1C8
CURRENT APPLICATION NUMBER: US/10/033,326
PRIOR FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: 60/095,325
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/112,851
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,145
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,511
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/115,558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119,341
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119,537
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/119,965
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: 60/162,506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170,262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/187,202
PRIOR FILING DATE: 2000-03-03

```

```

PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/14941
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
NUMBER OF SEQ ID NOS: 38
SEQ ID NO 22
LENGTH: 1029
TYPE: PRT
ORGANISM: Homo sapiens
US-10-033-326-22

Query Match 100.0%; Score 1029; DB 12; Length 1029;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHILVYHAMVILLTIGPPRADSEFOALDIWPEEKPPPTAFVDTSEALLPDMKL 60
Db 1 MHILVYHAMVILLTIGPPRADSEFOALDIWPEEKPPPTAFVDTSEALLPDMKL 60
Qy 61 RMIRSEVRLVDAIODELPOLLLEFVSGFIPVSSMSKILQFLDQVAHADPOTLEQNTM 120
Db 61 RMIRSEVRLVDAIODELPOLLLEFVSGFIPVSSMSKILQFLDQVAHADPOTLEQNTM 120
Qy 121 DKNYMAHLVEVHERGASGGQTFHSLTASIPRRRSTAPRKSSPEPIGGRIYVGT 180
Db 121 DKNYMAHLVEVHERGASGGQTFHSLTASIPRRRSTAPRKSSPEPIGGRIYVGT 180
Qy 181 QLRVLGPEDDLAGMFLQIPIPLSPDPWQSSPPRVALLQALGQELARVQSSPEVPGI 240
Db 181 QLRVLGPEDDLAGMFLQIPIPLSPDPWQSSPPRVALLQALGQELARVQSSPEVPGI 240
Qy 241 TVRVLOALATLLSSPHGALVMSMRSHPFLACPLRLQLCQYQRCVQDGTGSSFLKVL 300
Db 241 TVRVLOALATLLSSPHGALVMSMRSHPFLACPLRLQLCQYQRCVQDGTGSSFLKVL 300
Qy 301 OMLOWIDSPGVGGPRLAQLBMLASQASAGRRISDVRRGGLRLAEALAFRODLEVYSSIV 360
Db 301 OMLOWIDSPGVGGPRLAQLBMLASQASAGRRISDVRRGGLRLAEALAFRODLEVYSSIV 360
Qy 361 RAVIATLRSGEQSVEDLISKVQLIEVRSFHEBELLTAFPSATADAAPPACKPVY 420
Db 361 RAVIATLRSGEQSVEDLISKVQLIEVRSFHEBELLTAFPSATADAAPPACKPVY 420
Qy 421 VVSSLLLOEBEPLAGKPGADGSLAVRLGSSGLVDMLEMDPEVYSSCPDLOLRLL 480
Db 421 VVSSLLLOEBEPLAGKPGADGSLAVRLGSSGLVDMLEMDPEVYSSCPDLOLRLL 480
Qy 481 FSRKRGKGAQVPSFRPYLLTFTTHOSSWPTLHQICRVLLGKRRRQPSASLDPLMC 540
Db 481 FSRKRGKGAQVPSFRPYLLTFTTHOSSWPTLHQICRVLLGKRRRQPSASLDPLMC 540
Qy 541 IHVPRIMOGRDQTPQKRREELVLRVQGPGLSLVELLAEATRSQDDDTAACSILQAR 600
Db 541 IHVPRIMOGRDQTPQKRREELVLRVQGPGLSLVELLAEATRSQDDDTAACSILQAR 600
Qy 601 LPLLSCCGGDSEVRRVTEHLSGCIQMGDSVLRRCRDLLQLYLRBELRVPEVY 660
Db 601 LPLLSCCGGDSEVRRVTEHLSGCIQMGDSVLRRCRDLLQLYLRBELRVPEVY 660

```

Db 601 LPLLSCCGDDSVKRYTEHLSCG10QWGDVLRRCRDLLQLYLQRELPVPEVL 660  
QY 661 LHSEGAASSVCKLDGIHRFTLLADTSDSRALENRGADASMACKRLAVAHPLLLRHL 720  
Db 661 LHSEGAASSVCKLDGIHRFTLLADTSDSRALENRGADASMACKRLAVAHPLLLRHL 720  
QY 721 PMIAALLHGRTHLNFQEFROQNHLSCLFHLVGLLELLQPHVFRSEHOGALMDCLLSFIRL 780  
Db 721 PMIAALLHGRTHLNFQEFROQNHLSCLFHLVGLLELLQPHVFRSEHOGALMDCLLSFIRL 780  
QY 781 LLMVRRSSRLAAFINKEVQFIHKYITYNAPAAISFLQKHADPLHDLSFQNSDLVMLKSL 840  
Db 781 LLMVRRSSRLAAFINKEVQFIHKYITYNAPAAISFLQKHADPLHDLSFQNSDLVMLKSL 840  
QY 841 LAGLSLPSRDRDTRDGLDEGESESSAGSLPLVSVSLFTLTAEMAPYMKRLRSQGTVE 900  
Db 841 LAGLSLPSRDRDTRDGLDEGESESSAGSLPLVSVSLFTLTAEMAPYMKRLRSQGTVE 900  
QY 901 DLLEVLSDIDEMSRREPEILSFFSTNLQRLMSSAECCRNLAFLSRMNSPSIAAAF 960  
Db 901 DLLEVLSDIDEMSRREPEILSFFSTNLQRLMSSAECCRNLAFLSRMNSPSIAAAF 960  
QY 961 LPTFMVCLGSDQFEVYQVOTALRNLPEVALLCOEHAAVLLHRAFLVGMVGMQDPSAOISEAL 1020  
Db 961 LPTFMVCLGSDQFEVYQVOTALRNLPEVALLCOEHAAVLLHRAFLVGMVGMQDPSAOISEAL 1020  
QY 1021 RILHMEAVM 1029  
Db 1021 RILHMEAVM 1029

## RESULT 13

US-09-984-271-235  
; Sequence 235, Application US/09984271  
; Publication No. US20030040088A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 71 Human Secreted Proteins  
; FILE REFERENCE: P2030P1  
; CURRENT APPLICATION NUMBER: US/09/984, 271  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 09/482, 273  
; PRIOR FILING DATE: 2000-01-13  
; PRIOR APPLICATION NUMBER: PCT/US99/15849  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: 60/092, 921  
; PRIOR FILING DATE: 1998-07-15  
; PRIOR APPLICATION NUMBER: 60/092, 922  
; PRIOR FILING DATE: 1998-07-15  
; PRIOR APPLICATION NUMBER: 60/092, 956  
; PRIOR FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 267  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 235  
; LENGTH: 404  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-271-235

Query Match 39.3%; Score 404; DB 9; Length 404;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 IQQMGDSVLRRCRDLLQLYLQRELPVPEVLHSEGAASSVCKLDGIHRFTLL 685  
Db 1 IQQMGDSVLRRCRDLLQLYLQRELPVPEVLHSEGAASSVCKLDGIHRFTLL 60  
QY 686 ADSDSRALENRGADASMACKRLAVAHPLLLHRLPMAIALHGRTHLNFQEFROQNHLS 745  
Db 61 ADSDSRALENRGADASMACKRLAVAHPLLLHRLPMAIALHGRTHLNFQEFROQNHLS 120  
QY 746 CFLHVLGLLELLQPHVFRSEHOGALMDCLLSFIRLLNRYKSSRLAAFINKEVQFIHKY 805  
Db 746 CFLHVLGLLELLQPHVFRSEHOGALMDCLLSFIRLLNRYKSSRLAAFINKEVQFIHKY 805

Db 121 CFLHVLGLLELLQPHVFRSEHOGALMDCLLSFIRLLNRYKSSRLAAFINKEVQFIHKY 180  
QY 806 ITYNAPAAISFLQKHADPLHDLSFQNSDLVMLKSLAGLSLPSRDRDTRDGLDEGESESS 865  
Db 181 ITYNAPAAISFLQKHADPLHDLSFQNSDLVMLKSLAGLSLPSRDRDTRDGLDEGESESS 240  
QY 866 SAGSLPLVSVSLFTLTAEMAPYMKRLRSQGTVEEDLLEVLSDIDEMSRREPEILSFFST 925  
Db 241 SAGSLPLVSVSLFTLTAEMAPYMKRLRSQGTVEEDLLEVLSDIDEMSRREPEILSFFST 300  
QY 926 NLQRLMSSAECCRNLAFLSRMNSPSIAAAFLEPTFMVCLGSDQFEVYQVOTALRNLPE 985  
Db 301 NLQRLMSSAECCRNLAFLSRMNSPSIAAAFLEPTFMVCLGSDQFEVYQVOTALRNLPE 360  
QY 986 YALLCOEHAAVLLHRAFLVGMVGMQDPSAOISEALRILHMEAVM 1029  
Db 361 YALLCOEHAAVLLHRAFLVGMVGMQDPSAOISEALRILHMEAVM 404

## RESULT 14

US-09-984-271-148  
; Sequence 148, Application US/09984271  
; Publication No. US20030040088A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 71 Human Secreted Proteins  
; FILE REFERENCE: P2030P1  
; CURRENT APPLICATION NUMBER: US/09/984, 271  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 09/482, 273  
; PRIOR FILING DATE: 2000-01-13  
; PRIOR APPLICATION NUMBER: PCT/US99/15849  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: 60/092, 921  
; PRIOR FILING DATE: 1998-07-15  
; PRIOR APPLICATION NUMBER: 60/092, 922  
; PRIOR FILING DATE: 1998-07-15  
; PRIOR APPLICATION NUMBER: 60/092, 956  
; PRIOR FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 267  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 148  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (328)  
; OTHER INFORMATION: Xaa equals stop translation  
US-09-984-271-148

Query Match 31.8%; Score 327; DB 9; Length 328;  
Best Local Similarity 100.0%; Pred. No. 1; 8e-286;  
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 MACRKLAVAHPLLLHRLPMAIALHGRTHLNFQEFROQNHLSCLFHLVGLLELLQPHV 762  
Db 1 MACRKLAVAHPLLLHRLPMAIALHGRTHLNFQEFROQNHLSCLFHLVGLLELLQPHV 60  
QY 763 RSEHOGALMDCLLSFIRLLNRYKSSRLAAFINKEVQFIHKYITYNAPAAISFLQKHAD 822  
Db 61 RSEHOGALMDCLLSFIRLLNRYKSSRLAAFINKEVQFIHKYITYNAPAAISFLQKHAD 120  
QY 823 PLHDLSFQNSDLVMLKSLAGLSLPSRDRDTRDGLDEGESESSAGSLPLVSVSLFTPLT 882  
Db 121 PLHDLSFQNSDLVMLKSLAGLSLPSRDRDTRDGLDEGESESSAGSLPLVSVSLFTPLT 180  
QY 883 AAEMAPYMKRLRSQGTVEEDLLEVLSDIDEMSRREPEILSFFSTNLQRLMSSAECCRNLA 942  
Db 181 AAEMAPYMKRLRSQGTVEEDLLEVLSDIDEMSRREPEILSFFSTNLQRLMSSAECCRNLA 240  
QY 943 FSLALRMONSPSIAAFLPTFMVCLGSDQFEVYQVOTALRNLPEVALLCOEHAAVLLHRA 1002  
Db 943 FSLALRMONSPSIAAFLPTFMVCLGSDQFEVYQVOTALRNLPEVALLCOEHAAVLLHRA 1002

DB 241 FSLAIRSMQNSPSIAAFLPTMYCIGSODFEVOTALNLPYALLCOEHAAVLLHRAF 300  
QY 1003 LVGMYGOMDPSAQISEALRIHMEAVM 1029  
|||||  
DB 301 LVGMYGOMDPSAQISEALRIHMEAVM 327

Job time : 34 secs

RESULT 15  
US-09-764-891-4155  
; Sequence 4155, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIORITY FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4155  
; LENGTH: 161  
; TYPE: prt  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (40)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (84)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (94)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (96)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (98)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (105)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (136)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (137)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (139)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (141)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (143)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (145)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-4155

Query Match 4.2% Score 43; DB 9; Length 161;  
Best Local Similarity 100.0%; Pred. No. 6.1e-31;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 851 DRTDRGLDEGEERSSAGSLPLVSVSLFTPLTAEMAPYMKRL 893  
|||||  
DB 41 DRTDRGLDEGEERSSAGSLPLVSVSLFTPLTAEMAPYMKRL 83

Search completed: May 7, 2003, 19:47:23



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:42:59 ; Search time 48 Seconds  
(without alignments)  
4417.135 Million cell updates/sec

Title: US-09-929-769-7

Perfect score: 1029

Sequence: 1 MHILVHAMVILTLTGPPRA.....MDPSAISEALRIHMEAVM 1029

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database:

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	928	90.2	1377	4	Q9P2A8
2	809	78.6	812	4	Q96D36
3	698	67.8	698	4	Q8WV40
4	633	61.5	640	4	Q9Y3W8
5	260	25.3	673	4	Q9B7R1
6	100	9.7	310	4	Q9N7D1
7	36	3.5	818	11	Q91201
8	18	1.7	162	11	Q9CTF7
9	9	0.9	241	16	Q9KZM1
10	9	0.9	327	16	Q91345
11	8	0.8	81	3	Q9Y7G3
12	8	0.8	151	4	Q9UG55
13	8	0.8	154	4	Q9F9V0
14	8	0.8	157	5	Q9X7Y5
15	8	0.8	207	5	Q9VRE4
16	8	0.8	226	2	Q8VWZ8

17	8	0.8	249	2	Q8VR19	Q8VR19 myxococcus
18	8	0.8	253	11	Q54831	Q54831 mus musculus
19	8	0.8	267	11	Q35333	Q35333 mus musculus
20	8	0.8	272	17	Q972T6	Q972T6 sulfolobus
21	8	0.8	279	16	Q37R19	Q37R19 streptococ
22	8	0.8	311	5	Q9W4H0	Q9W4H0 dirosophila
23	8	0.8	320	16	Q8XW4	Q8XW4 raietonia s
24	8	0.8	345	5	Q44916	Q44916 caenorhabdi
25	8	0.8	379	17	Q8ZVJ6	Q8ZVJ6 pyrobaculo
26	8	0.8	386	16	Q99WL9	Q99WL9 staphylococ
27	8	0.8	423	16	Q33313	Q33313 mycobacteri
28	8	0.8	436	5	Q95T23	Q95T23 dirosophila
29	8	0.8	490	10	Q94CS4	Q94CS4 oryza sativ
30	8	0.8	519	2	Q8RP20	Q8RP20 rhodovulm
31	8	0.8	543	2	Q8RMJ1	Q8RMJ1 leptospira
32	8	0.8	552	4	Q8WU59	Q8WU59 homo sapien
33	8	0.8	593	16	Q8Z1W5	Q8Z1W5 yersinia pe
34	8	0.8	640	16	Q9RX24	Q9RX24 delinococcus
35	8	0.8	672	6	Q9BE88	Q9BE88 macaca fasc
36	8	0.8	677	11	Q8R251	Q8R251 mus musculu
37	8	0.8	705	4	Q9H9S2	Q9H9S2 homo sapien
38	8	0.8	728	2	Q9RH22	Q9RH22 allcyclobac
39	8	0.8	780	13	Q9YGM3	Q9YGM3 fugu rubrip
40	8	0.8	780	13	Q9PWQ1	Q9PWQ1 fugu rubrip
41	8	0.8	789	5	Q9H8F2	Q9H8F2 trypanosoma
42	8	0.8	853	4	Q9H8M5	Q9H8M5 homo sapien
43	8	0.8	882	4	Q9BOM8	Q9BOM8 homo sapien
44	8	0.8	893	4	Q9NDM1	Q9NDM1 homo sapien
45	8	0.8	913	17	Q8TVS4	Q8TVS4 methanopyru

## ALIGNMENTS

## RESULT 1

ID	Q9P2A8	PRELIMINARY:	PRT:	1377	AA.
AC	Q9P2A8:				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)				
DE	KIAA1440 protein (Fragment).				
GN	KIAA1440.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BRIN;				
RX	MEDLINE=20181126; PubMed=10718198;				
RA	Nagase T., Kikuno R., Ishikawa K., Hirotsawa M., Ohara O.				
RT	"Prediction of the coding sequences of unidentified human genes.XVI.				
RT	The complete sequences of 150 new cDNA clones from brain which code				
RL	for large proteins in vitro."				
DR	EMBL; AB037861; BAA92678.1;				
FT	NON_TER				
FT	SEQUENCE				
SO	1377	AA:	153070	MM:	CE70589EF594FE7E CRC64;
Query Match					
Best local similarity 99.9%; Pred. No. 0;					
Matches 1028; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	1	MHILVHAMVILTLTGPPRADSEFQALDIWPEKRPPTAFLVDTSEALLPWLKL	60		
DB	349	MHILVHAMVILTLTGPPRADSEFQALDIWPEKRPPTAFLVDTSEALLPWLKL	408		
QY	61	RMIRSEVLRLVPAALDLEPQQLLFVQSGRIVVSSMKLQFLDQAVAHDPOTLEONIM	120		
DB	409	RMIRSEVLRLVPAALDLEPQQLLFVQSGRIVVSSMKLQFLDQAVAHDPOTLEONIM	468		
QY	121	DKRYMAHLEVOHERGASGGQTFHSILTLASLPFRDSTAPPKSSPEQPIGGRIRVGT	180		

```

Db      469 DKNYMAHLYVQHERRGSSGQTFHSLTASLPKPRDSTEAPKPPSSPEQPIGQRIIVGT 528
QY      181 QLRVLPEDDLAQMFLQIFPLSPDPRMOWSSPPRVATLQOALQOELARVVOGPEVGI 240
Db      529 QLRVLPEDDLAQMFLQIFPLSPDPRMOWSSPPRVATLQOALQOELARVVOGPEVGI 588
QY      241 TVRYLQALATLSSPHGALVMSMHRSHFLACPLLRQCYQRCVPDGTGSSFLKVL 300
Db      589 TVRYLQALATLSSPHGALVMSMHRSHFLACPLLRQCYQRCVPDGTGSSFLKVL 648
QY      301 QMLOWIDSPVEGGPLRAOLRLMASQASAGRRLSDVRCGLRLAEALAFQDLEVSSTV 360
Db      649 QMLOWIDSPVEGGPLRAOLRLMASQASAGRRLSDVRCGLRLAEALAFQDLEVSSTV 708
QY      361 RAVIATLRSGEQCSVEPDLISKVLOGLIEVRSPHLEELLAFATADAAPPACKPV 420
Db      709 RAVIATLRSGEQCSVEPDLISKVLOGLIEVRSPHLEELLAFATADAAPPACKPV 768
QY      421 VVSSLLOEEEPPLAGKPGADGSLAVRLGSSGLLVDMLEMLDPEVSSCPDLQLRL 480
Db      769 VVSSLLOEEEPPLAGKPGADGSLAVRLGSSGLLVDMLEMLDPEVSSCPDLQLRL 828
QY      481 FSRKKGQAOVSPFRYLLTLTFHOSWPTLHOCIRVLLGKSPREOFDSSASIDELMAC 540
Db      829 FSRKKGQAOVSPFRYLLTLTFHOSWPTLHOCIRVLLGKSPREOFDSSASIDELMAC 888
QY      541 IHVPRIMOGDQRTPOKRRRELYLRVGPPELLISVELILAEATRSDDGDTAACSLIQAR 600
Db      889 IHVPRIMOGDQRTPOKRRRELYLRVGPPELLISVELILAEATRSDDGDTAACSLIQAR 948
QY      601 LPLLLSCCGDDSVKRVYTHLSGCIQOMGDSVYGRRCRDLLOLYQRELYRVPVEVL 660
Db      949 LPLLLSCCGDDSVKRVYTHLSGCIQOMGDSVYGRRCRDLLOLYQRELYRVPVEVL 1008
QY      661 LHSEGAASSVCKLDGLIHRFTLLADTSDSRALENRGADASMACKRLAVALPILLRL 720
Db      1009 LHSEGAASSVCKLDGLIHRFTLLADTSDSRALENRGADASMACKRLAVALPILLRL 1068
QY      721 PMIAALLHGRTHLNFQFROOHNLSCLHVLGLLELLQPHVFRSEHOGALMDCILSFIRL 780
Db      1069 PMIAALLHGRTHLNFQFROOHNLSCLHVLGLLELLQPHVFRSEHOGALMDCILSFIRL 1128
QY      781 LMYRKRSSRLAFAINKFVQFIKYYTYNAPAIISFLQKADPDLHDSFQNSDLVMLKSL 840
Db      1129 LMYRKRSSRLAFAINKFVQFIKYYTYNAPAIISFLQKADPDLHDSFQNSDLVMLKSL 1188
QY      841 LAGLSLPSRDDRTDRGLDEGESEESSAGSLPLVSVSLFTPLTAEMAPYMKRSLRGQTV 900
Db      1189 LAGLSLPSRDDRTDRGLDEGESEESSAGSLPLVSVSLFTPLTAEMAPYMKRSLRGQTV 1248
QY      901 DLLEVLSDIDEMSRRPETLSFSTNIQRLMSSAECCRNILAFSLARSMQNSPSIAAF 960
Db      1249 DLLEVLSDIDEMSRRPETLSFSTNIQRLMSSAECCRNILAFSLARSMQNSPSIAAF 1308
QY      961 LPTFMVCLGSDQDEEVYOTALRNLPEYALLQDEHAVALLHRAFLVGMGQMDPSAQISEAL 1020
Db      1309 LPTFMVCLGSDQDEEVYOTALRNLPEYALLQDEHAVALLHRAFLVGMGQMDPSAQISEAL 1368
QY      1021 RILHMEAVM 1029
Db      1369 RILHMEAVM 1377

```

```

OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LYMPH;
RA      Strausberg R.;
RL      Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC013367; AAH13367.1;
KW      Hypothetical protein.
FT      NON TER
SQ      SEQUENCE 812 AA; 89996 MW; 5054AE099C15BFB4 CRC64;

Query Match      78.6%; Score 809; DB 4; Length 812;
Best local Similarity 100.0%; Pred. No. 0;
Matches 809; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      221 QALQOELARVVOGPEVPGITVRYLQALATLSSPHGALVMSMHRSHFLACPLLRQCY 280
Db      4 QALQOELARVVOGPEVPGITVRYLQALATLSSPHGALVMSMHRSHFLACPLLRQCY 63
QY      281 YQRCVPDGTGSSFLKVLQMLQMLDPSVEGGPLRAOLRLMASQASAGRRLSDVRCGL 340
Db      64 YQRCVPDGTGSSFLKVLQMLQMLDPSVEGGPLRAOLRLMASQASAGRRLSDVRCGL 123
QY      341 LRLAEALAFQDLEVSSTVRAVATLRSGEQCSVEPDLISKVLOGLIEVRSPHLEELL 400
Db      124 LRLAEALAFQDLEVSSTVRAVATLRSGEQCSVEPDLISKVLOGLIEVRSPHLEELL 183
QY      401 AFSATADAAPPACKPVVVVSSLLQEEEPPLAGKPGADGSLAVRLGSSGLLVDM 460
Db      184 AFSATADAAPPACKPVVVVSSLLQEEEPPLAGKPGADGSLAVRLGSSGLLVDM 243
QY      461 LEMLDPEVSSCPDLQLRLFLFSRRKKGQAOVSPFRYLLTLTFHOSWPTLHOCIRVLL 520
Db      244 LEMLDPEVSSCPDLQLRLFLFSRRKKGQAOVSPFRYLLTLTFHOSWPTLHOCIRVLL 303
QY      521 GKSREQFDPDSASIDELMACIHVPRIMOGDQRTPOKRRRELYLRVGPPELLISVELILA 580
Db      304 GKSREQFDPDSASIDELMACIHVPRIMOGDQRTPOKRRRELYLRVGPPELLISVELILA 363
QY      581 EAETRSDDGDTAACSLIQARPLLLSCCGDDSVKRVYTHLSGCIQOMGDSVYGRRCRD 640
Db      364 EAETRSDDGDTAACSLIQARPLLLSCCGDDSVKRVYTHLSGCIQOMGDSVYGRRCRD 423
QY      641 LLLQLYQRELYRVPVEVLHSEGAASSVCKLDGLIHRFTLLADTSDSRALENRGAD 700
Db      424 LLLQLYQRELYRVPVEVLHSEGAASSVCKLDGLIHRFTLLADTSDSRALENRGAD 483
QY      701 ASMAKRLAVALPILLRLPMIAALLHGRTHLNFQFROOHNLSCLHVLGLLELLQPH 760
Db      484 ASMAKRLAVALPILLRLPMIAALLHGRTHLNFQFROOHNLSCLHVLGLLELLQPH 543
QY      761 VFRSEHOGALMDCILSFIRLLMYRKRSSRLAFAINKFVQFIKYYTYNAPAIISFLQKH 820
Db      544 VFRSEHOGALMDCILSFIRLLMYRKRSSRLAFAINKFVQFIKYYTYNAPAIISFLQKH 603
QY      821 ADPLHDSFQNSDLVMLKSLAGLSLPSRDDRTDRGLDEGESEESSAGSLPLVSVSLFT 880
Db      604 ADPLHDSFQNSDLVMLKSLAGLSLPSRDDRTDRGLDEGESEESSAGSLPLVSVSLFT 663
QY      881 LTAEMAPYMKRSLRGQTVDDLLEVLSDIDEMSRRPETLSFSTNIQRLMSSAECCRN 940
Db      664 LTAEMAPYMKRSLRGQTVDDLLEVLSDIDEMSRRPETLSFSTNIQRLMSSAECCRN 723
QY      941 LAFSLARSMQNSPSIAAALPTFMVCLGSDQDEEVYOTALRNLPEYALLQDEHAVALLHR 1000
Db      724 LAFSLARSMQNSPSIAAALPTFMVCLGSDQDEEVYOTALRNLPEYALLQDEHAVALLHR 783
QY      1001 AFLVGMGQMDPSAQISEALRILHMEAVM 1029
Db      784 AFLVGMGQMDPSAQISEALRILHMEAVM 812

```



RESULT 3  
08WV40 PRELIMINARY: PRT: 698 AA.  
AC 08WV40:  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypothetical 77.7 kDa protein (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SKIN;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC018777; AAH18777.1;  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 698 AA; 77709 MW; C917828091BF4778 CRC64;

Query Match 67.8%; Score 698; DB 4; Length 698;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 RLSDVGGGLRLAELARODLEVVSVTVRAVIAFLRSGECSVEPDLISKVLQGLIEVR 391  
DB 1 RLSDVGGGLRLAELARODLEVVSVTVRAVIAFLRSGECSVEPDLISKVLQGLIEVR 60  
QY 392 SPHEELLTAFFSATADASPFPACKPVVVVSSLLQEEEPAGKPGADGSLFAVRLG 451  
DB 61 SPHEELLTAFFSATADASPFPACKPVVVVSSLLQEEEPAGKPGADGSLFAVRLG 120  
QY 452 PSSGLVDMLEMDPEVSSCPDQLRLFLFSRRKKGQAQVPSFRPYLLTLFTHOSSWPT 511  
DB 121 PSSGLVDMLEMDPEVSSCPDQLRLFLFSRRKKGQAQVPSFRPYLLTLFTHOSSWPT 180  
QY 512 LHQCRVLLGKSREORPDSASIDFLMACIHVPRIWQGRDQTPQKREELVLRVQGPRL 571  
DB 181 LHQCRVLLGKSREORPDSASIDFLMACIHVPRIWQGRDQTPQKREELVLRVQGPRL 240  
QY 572 ISIVELIAEAETRSQDQDFAACSLIQARLPLLLSCCGDDESVRKVTEHLSGCIQOMGD 631  
DB 241 ISIVELIAEAETRSQDQDFAACSLIQARLPLLLSCCGDDESVRKVTEHLSGCIQOMGD 300  
QY 632 SVLGRCDLLQLQYLRPELRVPEVLLHSEGAASSVCKLDGLIHRFTLLADTSDS 691  
DB 301 SVLGRCDLLQLQYLRPELRVPEVLLHSEGAASSVCKLDGLIHRFTLLADTSDS 360  
QY 692 RALERGADSMACRKLAVAHPLLLRLPMTAALLHGRTHLNPOEFQONHLSGFLHVL 751  
DB 361 RALERGADSMACRKLAVAHPLLLRLPMTAALLHGRTHLNPOEFQONHLSGFLHVL 420  
QY 752 GULELLQPHVFRSEHOGALMDCLSFIRLLNLYRKSSRHIAAFINKFVOFIHKYITTYNAP 811  
DB 421 GULELLQPHVFRSEHOGALMDCLSFIRLLNLYRKSSRHIAAFINKFVOFIHKYITTYNAP 480  
QY 812 AAISFLQKHADPLHDLSFDNSDLVMLKSLAGLSLPSRDRTDRLDEEGEESASGLP 871  
DB 481 AAISFLQKHADPLHDLSFDNSDLVMLKSLAGLSLPSRDRTDRLDEEGEESASGLP 540  
QY 872 LVVSILFPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRRPILSFSTNLQRIAM 931  
DB 541 LVVSILFPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRRPILSFSTNLQRIAM 600  
QY 932 SSAECCCNLAFAFSLALRSMONSPSIAAFLPTFMYCLGSODEFVVQATALNLPETALLCO 991  
DB 601 SSAECCCNLAFAFSLALRSMONSPSIAAFLPTFMYCLGSODEFVVQATALNLPETALLCO 660  
QY 992 EHAVALLHRAFLVGMYGOMDPSAQSIALRLIHMEAVM 1029  
DB 660 EHAVALLHRAFLVGMYGOMDPSAQSIALRLIHMEAVM 1029

DB 661 EHAVALLHRAFLVGMYGOMDPSAQSIALRLIHMEAVM 698

RESULT 4  
09Y3W8 PRELIMINARY: PRT: 640 AA.  
AC 09Y3W8:  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Hypothetical 71.4 kDa protein (Fragment).  
OS Homo sapiens (Human).  
GN DKFZP586J0619.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=UTERUS;  
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL050110; CAB43278.1;  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 640 AA; 71375 MW; 1AC87FC2075B21A CRC64;

Query Match 61.5%; Score 633; DB 4; Length 640;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 ELTTAFESATADASPFPACKPVVVVSSLLQEEEPAGKPGADGSLFAVRLGSSGL 456  
DB 8 ELTTAFESATADASPFPACKPVVVVSSLLQEEEPAGKPGADGSLFAVRLGSSGL 67  
QY 457 LVDWLEMDLPEVSSCPDQLRLFLFSRRKKGQAQVPSFRPYLLTLFTHOSSWPTLHQCI 516  
DB 68 LVDWLEMDLPEVSSCPDQLRLFLFSRRKKGQAQVPSFRPYLLTLFTHOSSWPTLHQCI 127  
QY 517 RVLTKSRREORPDSASIDFLMACIHVPRIWQGRDQTPQKREELVLRVQGPRL 576  
DB 128 RVLTKSRREORPDSASIDFLMACIHVPRIWQGRDQTPQKREELVLRVQGPRL 187  
QY 577 LTLAEETRSQDQDFAACSLIQARLPLLLSCCGDDESVRKVTEHLSGCIQOMGD 636  
DB 188 LTLAEETRSQDQDFAACSLIQARLPLLLSCCGDDESVRKVTEHLSGCIQOMGD 247  
QY 637 RCRDILLQLYLRPELRVPEVLLHSEGAASSVCKLDGLIHRFTLLADTSDSRALEN 696  
DB 248 RCRDILLQLYLRPELRVPEVLLHSEGAASSVCKLDGLIHRFTLLADTSDSRALEN 307  
QY 697 RGADSMACRKLAVAHPLLLRLPMTAALLHGRTHLNPOEFQONHLSGFLHVL 756  
DB 308 RGADSMACRKLAVAHPLLLRLPMTAALLHGRTHLNPOEFQONHLSGFLHVL 420  
QY 757 LQPHVFRSEHOGALMDCLSFIRLLNLYRKSSRHIAAFINKFVOFIHKYITTYNAP 816  
DB 368 LQPHVFRSEHOGALMDCLSFIRLLNLYRKSSRHIAAFINKFVOFIHKYITTYNAP 427  
QY 817 LQKHADPLHDLSFDNSDLVMLKSLAGLSLPSRDRTDRLDEEGEESASGLP 876  
DB 428 LQKHADPLHDLSFDNSDLVMLKSLAGLSLPSRDRTDRLDEEGEESASGLP 547  
QY 877 LFTPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRRPILSFSTNLQRIAM 936  
DB 548 LFTPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDEMSRRPILSFSTNLQRIAM 607  
QY 937 CCRNLAFSLALRSMONSPSIAAFLPTFMYCLGSODEFVVQATALNLPETALLCOEHA 996  
DB 607 CCRNLAFSLALRSMONSPSIAAFLPTFMYCLGSODEFVVQATALNLPETALLCOEHA 607  
QY 997 LTHRAFLVGMYGOMDPSAQSIALRLIHMEAVM 1029  
DB 608 LTHRAFLVGMYGOMDPSAQSIALRLIHMEAVM 640

```

RESULT 5
O9BT91
ID 09BT91 PRELIMINARY; PRT; 673 AA.
AC 09BT91:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE Hypothetical 75.1 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC04286; AA04286.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 673 AA; 75055 MW; 6CE17654F5BA3DE CRC64;

Query Match
Best Local Similarity 25.3%; Score 260; DB 4; Length 673;
Matches 360; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MHLVYHAWVILTLGPRADSEFOALDIMPEEKPPTAFVDTSEALLPMLKL 60
DB 304 MHLVYHAWVILTLGPRADSEFOALDIMPEEKPPTAFVDTSEALLPMLKL 363
OY 61 MTRSEVRLVDAALDLEFQQLLEVFQSGFVSSKSLQFLDQAVADHPOTLEONIM 120
DB 364 MTRSEVRLVDAALDLEFQQLLEVFQSGFVSSKSLQFLDQAVADHPOTLEONIM 423
OY 121 DKYVMAHVVOHERGSGGOTFHSULTASLPPRODTEAPKPKSSPEQPIGGGRIRVGT 180
DB 424 DKYVMAHVVOHERGSGGOTFHSULTASLPPRODTEAPKPKSSPEQPIGGGRIRVGT 483
OY 181 QLRVLGPEDDLAGMFLQIFPLSPDPKMOSSPPRVALALQALGOELARVQSSPEVPGI 240
DB 484 QLRVLGPEDDLAGMFLQIFPLSPDPKMOSSPPRVALALQALGOELARVQSSPEVPGI 543
OY 241 TVAVLQALATLSSPHGALVMSHNSHFLACPLRLQCYQRCVPODTGFSFLKVL 300
DB 544 TVAVLQALATLSSPHGALVMSHNSHFLACPLRLQCYQRCVPODTGFSFLKVL 603
OY 301 QMLQWLDSPVEGGPLRAOLRMLASOASAGRRLSDVAGGLRLAEALAFRODLEVVSSTV 360
DB 604 QMLQWLDSPVEGGPLRAOLRMLASOASAGRRLSDVAGGLRLAEALAFRODLEVVSSTV 663
OY 361 R 361
DB 664 R 664

```

```

RA Ansoorge W., Wirkner U., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137358; CAB70710.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 310 AA; 34728 MW; EDBA23F1DOC08521 CRC64;

Query Match
Best Local Similarity 9.7%; Score 100; DB 4; Length 310;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 844 LSLPSRDRDRGIDEGEESSAGSLPLVSVSLFPTLAEMAPYKRLSRGQVDDL 903
DB 89 LSLPSRDRDRGIDEGEESSAGSLPLVSVSLFPTLAEMAPYKRLSRGQVDDL 148
OY 904 EVLSDIDEMSRRPETLSFSTNLRMSAECCRNIAF 943
DB 149 EVLSDIDEMSRRPETLSFSTNLRMSAECCRNIAF 188

RESULT 7
O91201
ID 091201 PRELIMINARY; PRT; 818 AA.
AC 091201:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE Hypothetical 90.9 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010333; AAH10333.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 818 AA; 90949 MW; A5128C8692BB7663 CRC64;

Query Match
Best Local Similarity 3.5%; Score 36; DB 11; Length 818;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 529 DPSASDPLWACIHVRIMGRODRTPOKRRRELV 564
DB 318 DPSASDPLWACIHVRIMGRODRTPOKRRRELV 353

RESULT 8
O9CTF7
ID 09CTF7 PRELIMINARY; PRT; 162 AA.
AC 09CTF7:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE 1110015K06R1k protein (Fragment).
GN 1110015K06R1K
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shidata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,

```

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirral L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombers P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.",  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK003728; BAB22963.1; -  
 DR MGD: MGI:1915760; 1110015K06R1k.  
 FT NON\_TER 1  
 SO SEQUENCE 162 AA; 17939 MW; E343A7CF040CEC7A CRC64;

Query Match 1.7%; Score 18; DB 11; Length 162;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 988 LICEHAVALLHRAFLVG 1005  
 |||||  
 Db 121 LICEHAVALLHRAFLVG 138

RESULT 9  
 Q9RZNI PRELIMINARY; PRT; 241 AA.

AC Q9RZNI; 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE DNA-binding stress response protein, DPS family.  
 GN DRB0092.  
 OS Deinococcus radiodurans.  
 OC Plasmid Mpl.  
 CC Bacteria; Thermus/deinococcus group; Deinococci; Deinococcales;  
 CC Deinococcaceae; Deinococcus.  
 CX NCBI\_TaxID=1299;  
 OX [1]  
 RN SEQUENCE FROM N.A.

RP STRAIN-R1;  
 RC MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Motif K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vanatvean J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.,  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1.",  
 RL Science 286:1571-1577(1999).  
 DR EMBL: AE001826; AAF12541.1; -  
 DR TIGR: DRB0092; -  
 DR InterPro: IPR002177; DPS.  
 DR Pfam: PF02047; DPS; 1.  
 DR Prodom: PD149803; DPS; 1.  
 KW DNA-binding; Plasmid; Complete proteome.  
 SO SEQUENCE 241 AA; 26089 MW; 18F6C6F5D629E75 CRC64;

Query Match 0.9%; Score 9; DB 16; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 4.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 419 VVVVSSLL 427  
 |||||  
 Db 8 VVVVSSLL 16

RESULT 10  
 O91345 PRELIMINARY; PRT; 327 AA.

AC O91345; 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Hypothetical protein PA1680.  
 GN PA1680.  
 OS Pseudomonas aeruginosa.  
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.",  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004595; AAG05069.1; -  
 DR InterPro: IPR000734; Lipase.  
 DR InterPro: IPR000379; Ser\_estrs\_site.  
 DR PROSITE: PS00120; LIPASE\_SER; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 327 AA; 35210 MW; C3DD0BA9351A1E67 CRC64;

Query Match 0.9%; Score 9; DB 16; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 340 LRLRLAALA 348  
 |||||  
 Db 81 LRLRLAALA 89

RESULT 11  
 O9Y7G3 PRELIMINARY; PRT; 81 AA.

AC O9Y7G3; 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Cop c1 allergen (Fragment).  
 GN COP C1.  
 OS Coprinus comatus (Shaggy mane).  
 CC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 CC Agaricales; Agaricaceae; Coprinus.  
 CX NCBI\_TaxID=56187;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=CAP;  
 RX MEDLINE=99414166; PubMed=10482839;  
 RA Brander K.A., Bobely P., Cramer R., Pichler W.J., Heibling A.,  
 RT "IGE-binding proliferative responses and skin test reactivity to Cop c  
 RT 1, the first recombinant allergen from the basidiomycete Coprinus  
 RT comatus.",  
 RL J. Allergy Clin. Immunol. 104:630-636(1999).  
 DR EMBL: AJ132235; CAB39376.1; -  
 FT NON\_TER 1  
 SO SEQUENCE 81 AA; 8967 MW; E9CD4842DA24BB70 CRC64;

Query Match 0.8%; Score 8; DB 3; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 710 VAAPELLL 717

Db 35 VARPLRL 42

## RESULT 12

09UG55 PRELIMINARY; PRT; 151 AA.  
 AC 09UG55;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE Hypochemical 16.9 kDa protein (Fragment).  
 GN DKFZP434A091.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Poustika, A., Klein M., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL050372; CAB62530.1;  
 DR InterPro: IPR000120; Amidase.  
 DR Pfam: PF01425; Amidase; 1.  
 FT Hypochemical protein.  
 KW NON\_TER  
 SQ SEQUENCE 151 AA; 16894 MW; D16F7383BADAA4B4D CRC64;

## Query Match

Best Local Similarity 0.8%; Score 8; DB 4; Length 151;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 713 PLLLRL 720

Db 7 PLLLRL 14

## RESULT 13

09F9V0 PRELIMINARY; PRT; 154 AA.  
 AC 09F9V0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Paal.  
 GN PAAL.  
 OS Azococcus evansii.  
 OC Bacteria; Proteobacteria; beta subdivision; Rhodocyclus group;  
 OC Azococcus.  
 OX NCBI\_TaxID=59406;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KB740;  
 RA El-Said Mohamed M.;  
 RT "Biochemical and molecular characterization of phenylacetate-coenzyme  
 A ligase, an enzyme catalyzing the first step in aerobic metabolism of  
 RT phenylacetic acid in Azococcus evansii.";  
 RL J. Bacteriol. 182:286-294(2000).  
 DR EMBL: AF176259; AAC28967.1;  
 DR InterPro: IPR003736; DUF157.  
 DR Pfam: PF02584; DUF157; 1.  
 DR TIGRfams: TIGR00369; unchar\_dom\_1; 1.  
 SQ SEQUENCE 154 AA; 16550 MW; 030AC598B5A441C3 CRC64;

## Query Match

Best Local Similarity 0.8%; Score 8; DB 2; Length 154;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 681 FITLADT 688

Db 69 FITLADT 76

## RESULT 14

09XTY5 PRELIMINARY; PRT; 157 AA.  
 AC 09XTY5;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE T03D8.2 protein.  
 GN T03D8.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mortimore B.J.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "genome sequence of the nematode C. elegans: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z92838; CAB07406.1;  
 DR InterPro: IPR000230; Ribosomal\_S12.  
 DR Pfam: PF00164; Ribosomal\_S12; 1.  
 DR PRINTS: PR01034; Ribosomal\_S12.  
 DR PRODOM: PD000576; Ribosomal\_S12; 1.  
 DR TIGRfams: TIGR00961; TpsL\_dact; 1.  
 SQ SEQUENCE 157 AA; 17370 MW; 6C6A837282EB26C CRC64;

## Query Match

Best Local Similarity 0.8%; Score 8; DB 5; Length 157;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 TLFTHQSS 508

Db 27 TLFTHQSS 34

## RESULT 15

09VRE4 PRELIMINARY; PRT; 207 AA.  
 AC 09VRE4;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE Cg1739 protein.  
 GN Cg1739.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003569; AAF50857.1; -  
 DR FLYBase: FBgn0031152; CG1739.  
 SO SEQUENCE 207 AA; 23166 MM; 2ACBAC1BB60EA197 CRC64;

## Query Match

0.8%; Score 8; DB 5; Length 207;

Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 LQEEEPPLA 434  
 |||||||||  
 DB 145 LQEEEPPLA 152

Search completed: May 7, 2003, 19:45:51  
 Job time : 55 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 13, 2003, 21:52:01 ; Search time 525 Seconds

(without alignments)  
4413.917 Million cell updates/sec

Title: US-09-929-769-7

Sequence: 1 MHLLVHAMVILTLGPPRA.....MDPSAOISEALRIILHEAVM 1029

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ .p2n.model -DEV-xlh  
-O/cgq2\_1/USPTO.spool/US09929769/runat\_07052003\_161923\_8024/app\_query.fasta.1.1223  
-DB-nucleicseq.101002 -OPMT-fastlap -SUFFIX-p2n.rng -MINMATCH=0.1 -LOOPCT=0  
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALLIN=200 -THR.SCORE-pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE-LOCAL -OUTFMT-pco -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR-US09929769.ecgn.1.1.34.4.runat.07052003.161923.8024 -NCPU=6 -ICPU=3  
-NO.XLPXY -NO.MAP -LARGEJURY -NEG\_SCORES=0 -NAIT -LONGQS -DEV.TIMEOUT=120  
-MAIN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5241	100.0	3437	21 AAC58109	Human PRO3434 nucl
2	5241	100.0	3437	21 AAA51266	Human DNA encoding
3	5241	100.0	3437	24 ABK11743	DNA encoding tumou
4	4211	80.3	2824	24 ABK35138	Human cDNA encodin
5	3229	61.6	2275	21 AAC76704	Human ORFX ORF2259
6	2055	39.2	1539	21 AAZ98062	Human secreted pro
7	2055	39.2	1539	21 AAZ98062	Human secreted pro
8	2055	39.2	1539	21 ABK69771	Human secreted pro
9	924	17.6	639	20 AAV8653	EST clone C2496.
10	891	17.0	526	22 ABA08362	Human secreted pro
11	863	16.5	6207	23 ABLO5745	Drosophila melanog
12	863	16.5	8355	23 ABLO5744	Drosophila melanog
13	592	11.3	484	22 ABL01467	Human reproductive
14	228	4.4	4351	23 ABL07340	Drosophila melanog
15	193	3.7	119	21 AAC11139	Human secreted pro
16	163.5	3.1	14800	24 ABL66291	Lung cancer relate
17	157.5	3.0	7419	23 AAS51427	Pseudomonas aerugi
18	157	3.0	33529	23 AAS17367	DNA sequence of S.
19	156.5	3.0	49377	19 AAV05287	The soraphen biosy
20	153.5	2.9	14835	24 AAS94858	Human DNA sequence
21	152.5	2.9	4080	19 AAV34368	Human protein tyro
22	152.5	2.9	4080	20 AA211221	Human protein tyro
23	152.5	2.9	4080	24 ABL64464	Stomach cancer rel
24	151.5	2.9	3787	22 ABA01228	Putative rbp inter
25	151.5	2.9	6045	22 AAH77994	Nucleotide sequenc
26	151.5	2.9	7093	24 AAD38851	Human kinase (PKIN
27	151	2.9	3250	23 AAS84219	DNA encoding novel
28	150.5	2.9	28958	18 AAT89956	Sorangium cellulos
29	150.5	2.9	28958	21 AAT75299	DNA sequence of So
30	150	2.9	28598	17 AAT06765	Sorangium cellulos
31	148	2.9	15355	22 AAB57405	Human skeletal mus
32	148	2.8	15359	22 ABA08724	Human ryanodine re
33	148	2.8	15359	22 AAS69156	DNA encoding novel
34	148	2.8	15464	12 AA011415	Ryanodine receptor
35	147	2.8	4031	20 AAV69717	Tumour rejection a
36	147	2.8	4121	24 ABR83720	Human cDNA differe
37	147	2.8	4121	24 ABL62861	Breast cancer rela
38	147	2.8	4121	24 ABL63082	DNA encoding novel
39	147	2.8	5526	23 AAS76845	Human polynucleoti
40	147	2.8	5778	22 AAK53086	Human cDNA encodin
41	147	2.8	5804	22 AAK52102	Human ORFX ORF94
42	147	2.8	11736	22 AAK52652	Human ORFX ORF94
43	146.5	2.8	4224	21 AAC74939	Human chromosome 1
44	145.5	2.8	4827	22 AAH22143	Tumour rejection a
45	145	2.8	4225	20 AAV69720	

## ALIGNMENTS

RESULT 1  
AAC58109  
ID AAC58109 standard; cDNA: 3437 BP.

AC AAC58109;  
25-JAN-2001 (first entry)

DE Human PRO3434 nucleotide sequence SEQ ID NO:22.

KW Human: tumour; diagnosis; neoplastic disease; proliferation; cancer;

KW Identification; tumorigenesis; anticancer; detection; ss.

OS Homo sapiens.

XX WO200053750-A1.

XX 14-SEP-2000.

PCR/US99/28551

XX 02-DEC-1999; 99WO-US28551.  
 XX 08-MAR-1999; 99WO-US05028.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 01-DEC-1999; 99WO-US28634.  
 XX (GERTH ) GENENTECH INC.  
 PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WT;  
 DR WPI, 2000-594320/56.  
 P-PSDB: AAB24027.  
 XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit  
 PT the growth of tumors in mammals, and to identify inhibitors of PRO  
 PT polypeptide activity or expression -  
 PS Claim 50; Fig 13; 226pp; English.

CC The present invention describes an antibody that binds to a human  
 CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;  
 CC PRO434; PRO1927; PRO3567; PRO1293; PRO1303; PRO344; PRO4354;  
 CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has  
 CC anticancer activity and can be used to diagnose tumors in mammals, by  
 CC detecting complex formation when the antibody is contacted with test  
 CC cells. Increased expression of genes encoding (I) can also be detected  
 CC to diagnose tumors. Agents which inhibit the activity of (I),  
 CC especially the antibodies, or an antisense oligonucleotide which  
 CC hybridizes to genes encoding (I), can be used to inhibit tumor growth,  
 CC preferably by inducing cell death. Methods from the present invention  
 CC can be used to identify compounds which inhibit the biological activity  
 CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation  
 CC probes used in examples from the present invention for human PRO  
 CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human  
 CC PRO polynucleotide and protein sequences given in the exemplification of  
 CC the present invention.

CC Sequence 3437 BP; 558 A; 1186 C; 1098 G; 595 T; 0 other;

Alignment Scores:  
 Pred. No.: 0 Length: 3437  
 Score: 5241.00 Matches: 1029  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-929-769-7 (1-1029) x AAC58109 (1-3437)

QY 1 MethHisIleuValAlaHisAlaMetValIleLeuLeuThreLeuGlyProPArgAla 20  
 DB 46 ATGCACATCTCTGCTGCTCCATGCATGATGATCTCTGATGCGGCCCTCCAGCC 105  
 QY 21 AspAspSerGluPheGlnAlaLeuLeuAspIleTyrPheProGluGluLysProLeuPro 40  
 DB 106 GACGACAGGAGATTCCAGCGCTGACATCTGTTCCGAGAGAGAAGACCATGCGCC 165  
 QY 41 ThrAlaPheLeuValAspThrSerGluGluAlaLeuLeuProAspThrLeuLysLeu 60  
 DB 166 ACCGCTCTCTGCTGACACATCGAGAGAGCCCTCTCTCTCTACCTGCTGAAGCTG 225  
 QY 61 ArgMetIleArgSerGluValLeuArgLeuValAspAlaAlaLeuGlnAspLeuGluPro 80  
 DB 226 CGCATATCGCTTCTGAGGCTCGCCCTGCTGAGCGCCCTGCAGAGACTGAGCCG 285  
 QY 81 GlnGlnLeuLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerLysLeu 100  
 DB 286 CAGCAGCTCTCTGCTGCGAGTGTGAGCATCCCGCTGCCAGCATGAGCAAACTC 345  
 QY 101 LeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGluGlnAsnIleMet 120

DB 346 CTCACCTTCTCTGACAGCAGCGCCAGACGCCCACTCTGAGACGAACATCATG 405  
 QY 121 AspLysAsnTyrMetAlaHisLeuValGluValGlnHisGluArgGlyLysLeu 140  
 DB 406 GACAAAGATTACATGAGCCCACTGTGAGAGCTCCACATGAGCGCCGCTCCGAGGC 465  
 QY 141 GlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgArgAspSerThrGlnAla 160  
 DB 466 CAGACTTTCACCTCTGCTGACAGCTCCCTGCGCCCGCCGAGACAGACAGAGGCA 525  
 QY 161 ProLysProLysSerSerProGluGlnProIleGlyGlnGlyArgIleArgValGlyThr 180  
 DB 526 CCCAAACCAAGAGCAGCCAGCCAGCCATGAGCCAGCGGCGGATGCGGGGAGAC 585  
 QY 181 GlnLeuArgValLeuGluIleProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200  
 DB 586 CAGCTCCGGGTGCTGGCCCTGAGAGACCTGCGGATGTCTCTCCAGATTTCGCCG 645  
 QY 201 LeuSerProAspProArgTyrGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220  
 DB 646 CTCAGCCCGGAGCCCTGCGTGGAGAGCTCCAGTCCCGCCGCTGCGCCCTGCGAG 705  
 QY 221 GlnAlaLeuGlyGlnGlnLeuAlaArgValValGlnGlySerProGluValProGlyIle 240  
 DB 706 CAGGCCCTGGCCAGAGAGCTGGCCCTGCTCCAGAGGAGCCGAGGTGCGGGCATC 765  
 QY 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyGlyAlaLeu 260  
 DB 766 ACGGTGCTGTCTCGAGGCCCTCGCCACCTCTCTAGCTCCCAACGCGGTGCTG 825  
 QY 261 ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuCysGln 280  
 DB 826 GTGATGTTCATGACGCGTGAACCTTCCTGCGCCCTGCTGCGCCAGCTCTCCAG 885  
 QY 281 TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuLysValLeuLeu 300  
 DB 886 TACCAGCGCTGTGTGACAGGACCGGCTTCTCTGCTCTTCTGGAAGTGTCTCTG 945  
 QY 301 GlnMetLeuGlnThrLeuAspSerProGlyValGluGlyLysProLeuArgAlaGlnLeu 320  
 DB 946 CAGATGCTGAGTGGCTGAGCAGCCCTGGGCTGAGAGCGCGCCCTGCGGGCAGACGTC 1005  
 QY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLysSerAspValArgGlyLeu 340  
 DB 1006 AGCATGCTTGCAGCCAGCGCTTCCAGCGCGGAGCTCAGTATGTGAGAGGGGCTC 1065  
 QY 341 LeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValValSerSerThrVal 360  
 DB 1066 CTGCGCTGCGCAGGCTCTGCTGCTTCCGTCAAGACTGAGAGTGTCACTCACTGTC 1125  
 QY 361 ArgAlaValIleAlaThrLeuArgSerGlyGlnGlnCysSerValGluProAspLeuIle 380  
 DB 1126 CGTGCCGTATCGCCACCCCTGAGAGCTGAGGAGACAGTGCAGCGGAGCTGATC 1185  
 QY 381 SerLysValLeuGlnGlnLeuIleGluValArgSerProHisLeuGlnLeuLeuThr 400  
 DB 1186 AGCAAAAGTCTCCAGGCGCTGATGAGTGAAGTCCCTCCAGTGAAGAGAGCTGTACT 1245  
 QY 401 AlaPhePheSerAlaThrAlaAspAlaAlaSerProPheProAlaCysLysProValAla 420  
 DB 1246 GCATTCTCTCTCCACTGCGAGTGTCTCTCCCTTTCACACCTGTAAAGCCGTTGTC 1305  
 QY 421 ValValSerSerLeuLeuLeuGlnGlnGluGluProLeuAlaGlyLysProGlyAla 440  
 DB 1306 GTGGTAGCTCCCTGCTGCTGAGAGAGAGACCCCTGCTGGGGGAAAGCGGGTGGC 1365  
 QY 441 AspGlyGlySerLeuLeuAlaValArgLeuGlyProSerSerSerLysLeuValAspTrp 460  
 DB 1366 GACGTGTGGAGCTTGAAGGCGGTGCGGCGCTGCGAGGCTCTCTAGTGAAGCTG 1425  
 QY 461 LeuGlnMetLeuAspProGluValValSerSerCysProAspLeuGlnLeuArgLeuLeu 480





PR 12-JAN-1999; 9905-0115558.  
 PR 12-JAN-1999; 9905-0115565.  
 PR 12-JAN-1999; 9905-0115733.  
 PR 09-FEB-1999; 9905-0119341.  
 PR 10-FEB-1999; 9905-0119347.  
 PR 12-FEB-1999; 9905-0119365.  
 PR 02-JUN-1999; 9905-0122522.  
 XX (GETH ) GENENTECH INC.

PI Boistein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A,  
 PI Currey AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;  
 PI Wood WL;  
 XX WPI, 2000-431586/37.  
 DR P-PSDB; AAY96736.

XX Isolated nucleic acid molecule encodes a PRO polypeptide which is a  
 PT transmembrane polypeptide  
 XX Claim 12; Fig 15; 154pp; English.

CC This cDNA encodes PRO3434, a novel secreted protein. The invention  
 CC concerns novel secreted and transmembrane proteins, designated PRO  
 CC polypeptides. The cDNA and gene sequences are useful in the recombinant  
 CC production of PRO polypeptides, as a hybridization probe to screen  
 CC libraries to isolate cDNAs with sequence identity to PRO polypeptides or  
 CC to map the gene encoding the PRO polypeptides and analyzing genetic  
 CC disorders. The cDNA/gene can also be used to produce transgenic animals  
 CC useful for the development and screening of therapeutically useful  
 CC reagents. They can also be used in gene therapy, e.g. to replace a  
 CC defective gene.

XX Sequence 3437 BP; 558 A; 1186 C; 1098 G; 595 T; 0 other;

# Alignment Scores:

Pred. No.: 0 Length: 3437  
 Score: 5241.00 Matches: 1029  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-929-769-7 (1-1029) x AA51266 (1-3437)

QY 1 MethHisIleLeuValValHisAlaMetValIleLeuLeuThrLeuGlyProProArgAla 20  
 DB 46 ATGCACATCTCTGTCGTCATGCATGTCATCTCTGACGCTGGCGCCCTCGAGCC 105  
 QY 21 AsparSerGluPheGlnAlaLeuLeuAspIleTrrPheProGluGluProLeuPro 40  
 DB 106 GACGACAGCGAGTCCAGCGCTGCTGACATCTGTTCCGAGAGAGAGCCACTGCC 165  
 QY 41 ThrAlaPheLeuValAspTrpSerGluGluAlaLeuLeuProAspTrpLeuVal 60  
 DB 166 ACCGCTTCTCTGTCGACACATCGAGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 225  
 QY 61 ArgMetIleArgSerGluValLeuArgLeuValAspAlaAlaLeuGlnAspLeuPro 80  
 DB 226 CCATGATCCGTTCTGAGTGTCTCCCGCTGTCGAGCGCCCTCGAGAGCTCGAGCG 285  
 QY 81 GlnGlnIleLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerIleu 100  
 DB 286 CAGCAGCT 345  
 QY 101 LeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGluGlnAspIleMet 120  
 DB 346 CTCGACTTCTCTGACAGCAGCAGTGGCCACAGACCCCAACATCTGAGAGAGAAATCATG 405  
 QY 121 AspLysAsnTyrMetAlaHisLeuValGluValGlnHisGluArgGlyAlaSerGlyGly 140  
 DB 406 GACAGAAATTAATGATGGCCACCTGTCGAGTGTGACAGCATGAGGCGGCGCTCGAGGC 465

QY 141 GlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgArgAspSerThrGluAla 160  
 DB 466 CAGACTTTCACATCTCTGCTACAGCCTCCCTGCGCCCGCCGAGACAGCAGAGCA 525  
 QY 161 ProLysProLysSerSerProGluGlnProIleGlyGlnGlyArgIleArgValGlyThr 180  
 DB 526 CCCAAACCAAGAGCAGCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 585  
 QY 181 GlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200  
 DB 586 CAGCTCTGGGCTCTGCGCCCTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 645  
 QY 201 LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220  
 DB 646 CTCAGCCCGAGCCTCTGTCGAGAGCTCAGTCCCGCCCGCTGCTCTCTCTCTCTCTCTCT 705  
 QY 221 GlnAlaLeuGlyGlnGluLeuAlaArgValValGlnGlySerProGluValProGlyIle 240  
 DB 706 CAGGCTCTGGGCGAGAGCTGGCCCGCTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 765  
 QY 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyAlaLeu 260  
 DB 766 AGGTCGTCGTCCTGTCAGGCTCTGCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 825  
 QY 261 ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuCysGln 280  
 DB 826 GTGATGTCATCAGCCGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 885  
 QY 281 TyrGlnArgCysValProGlnAspTrpGlyPheSerSerLeuPheLeuValLeuLeu 300  
 DB 886 TACAGCCGTCGTGTGTCAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 945  
 QY 301 GlnMetLeuGlnTrpLeuAspSerProGlyValGluGlyProLeuArgAlaGlnLeu 320  
 DB 946 CAGATGCTCAGAGCTGTCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1005  
 QY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeu 340  
 DB 1006 AGGATCTTGGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1065  
 QY 341 LeuArgLeuAlaGluValaLeuAlaPheArgGlnAspLeuGluValValSerSerThrVal 360  
 DB 1066 CTGCGCTGGCGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125  
 QY 361 ArgAlaValIleAlaThrLeuArgSerGlyGlnCysSerValGlnProAspLeuIle 380  
 DB 1126 CGTCCGTCATGCCACCTGAGGTCGTGGGAGACATGTCAGCGTGGAGCCGCTGATC 1185  
 QY 381 SerLysValLeuGlnGlyLeuIleGluValArgSerProHisLeuGlnIleLeuLeuThr 400  
 DB 1186 AGCAAAATCTCCAGGGGCTGATCGAGTGAAGTCCCGCCACCTGAGAGACTCTGACT 1245  
 QY 401 AlaPheSerAlaThrAlaAspAlaAspProPheProAlaCysLysProValVal 420  
 DB 1246 GCATTTCTTCTGSCACTCGGATGCTGCTCCCGTTTCAAGCCGTGAAGCCGCTTGG 1305  
 QY 421 ValValSerSerLeuLeuGlnGluGluGluProLeuAlaGlyLysProGlyAla 440  
 DB 1306 GTGTGAGCTCTCTGCTGTCAG 1365  
 QY 441 AspGlyGlySerLeuGlnValAlaArgLeuGlyProSerSerGlyLeuLeuValAspTrp 460  
 DB 1366 GAGGTGGAGAGCTGAGAGCCTGCGGCTGGGGCTCTGTCAGGCTCTAGTGAATG 1425  
 QY 461 LeuGlnLeuLeuAspProGluValValSerSerCysProAspLeuGlnLeuArgLeuLeu 480  
 DB 1426 CTGGAATGCTTGAGCCGAGAGTGTGACAGCTGTCGCGAGCTGAGCTGAGCTGCTG 1485  
 QY 481 PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProGlyLeuLeu 500  
 DB 1486 TTCTCCCGGAGAGAGAGGCAAGGTCAAGGTCAGGCTCCCTGCTCTCTCTCTCTCTCTCT 1545  
 QY 501 ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeuLeu 520

Db	1546	ACCCTTCACAGCATGATGTCACACTGGCCACACTGCACACTGATCGAATCGAATCTGCTG	1605
Oy	521	GlyylsSerArgIuGlnIarGpheAspProSerAlaSerLeuAspPheLeuTrrPalAcys	540
Db	1606	GGCAAGAGCCGGGACAGAGATTGCAGACCCCTCGCTCTGAGACTTCTCGAGCCCTGC	1665
Oy	541	IleHisValProArgIleTrrPngIlyArgAspGlnArgThrProGlnLysArgArgIu	560
Db	1666	ATCCATGTCTCTCGCATCTGGCAGAGGGGGGAGACACCGACCCCGCAGAAAGCGCGGAG	1725
Oy	561	GluLeuValLeuArgValGInGInIyProGluLeuIleSerLeuValGluLeuIleLeuAla	580
Db	1726	GAGTGGTGCTGGGGGTCCAGGGGCCGGAGACTATAGCTGTGTGGACTGATCTGGCC	1785
Oy	581	GluAlaGluThrArgSerGlnAspGlyAspPrrAlaAlaCysSerLeuIleGlnAlaArg	600
Db	1786	GAGCGGAGAGCGGAGACCAGAGAGGGGAGACAGCGCGCTCAGGCTCATCAGACGCCCG	1845
Oy	601	LeuProLeuLeuLeuSerCysCysGlyAspAspGluSerValArgLysValIrrGlu	620
Db	1846	CTGGCCCTGGTGGCTCACACTCTCTGTGGGAGAGATGATGAGAGAGGTGCTGCAG	1905
Oy	621	HisLeuSerGlyCysIleGInGInTrrPglYAspSerValLeuGlyArgArgCysArgsp	640
Db	1906	CACCTGTACAGCTGTACATCCAGCAAGTGGGAGAGAGCTGTGGAAAGCGCTGCCCGAAC	1965
Oy	641	LeuLeuLeuGlnLeuTrrLeuGlnArgCProGluLeuArgValProValProGluValLeu	660
Db	1966	CTTCTCTCGAGAGCTTACATCCACAGCGCGCGAGACTCGGGGTCCCTGCTAGAGTCTTA	2025
Oy	661	LeuHisSerGluGlyAlaAlaSerSerSerValCysLysLeuAspGlyLeuIleHisArg	680
Db	2026	CTGCACAGCAAGGCGCTGCACACACACAGCTGTCAAGCTGCAGAGACTCATCCAGCC	2085
Oy	681	PheIleThrLeuLeuAlaAspPrrTrsAspSerArgAlaLeuGlnAsnArgGlyAlaAsp	700
Db	2086	TTTCATCACGCTCTGTGGGACACACAGCATCCCGGGCTGTGGAAACCGAGGGCGGAT	2145
Oy	701	AlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuLeuArgHisLeu	720
Db	2146	GCGAGCATGGCGCTGGCGGAGCTGGGGGTGGGCGACCCGCTCTCTGTCAGGACTGTG	2205
Oy	721	PrometIleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPheGlnIuPheArgGln	740
Db	2206	CCCATGTATCCGGGGCTCTGCACAGGCGCCACCCACTCAACTTCACAGAGATTCGGCAG	2265
Oy	741	GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuGlnProHis	760
Db	2266	CAGACACCACTGAGCTGTCTTCGACAGTGGGTGGGCTGTGGACTCTGCAGCGGAC	2325
Oy	761	ValPheArgSerGlnHisGInGlyAlaLeuTrrPspCysLeuLeuSerPheIleArgLeu	780
Db	2326	GTGTTCGGGACGAGACGACAGGGGGCGCTGTGGAGCTGCTTCTTCTATCCGCGCTG	2385
Oy	781	LeuLeuAsnTrrArgLysSerSerArgHisLeuAlaAlaPheIleAsnLysPheValGln	800
Db	2386	CTGTGTAATTACAGGAAGTCTCCCGCCCAATGTGGTGTCTTATCAACAAGTTGTGCAG	2445
Oy	801	PheIleHisLysTrrIleThrTrrAsnAlaProAlaAlaIleSerPheLeuGlnLysHis	820
Db	2446	TTTCATCATAGTACATTACTCATCAATGCCCGCAGCAGCCATCTCTTCTCAGAGAC	2505
Oy	821	AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu	840
Db	2506	GCGGACCCGCTCCACAGCTCTCTTTCGACAACAGTGACTGTGTATCTCAAAATCCCTC	2565
Oy	841	LeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArgGlyLeuAspGluGln	860
Db	2566	CTTCGAGGCTCAGCGCGCCACAGGAGACAGAGACCGAGGCGCTGACGAGAG	2625
Oy	861	GlyGluGlnIuSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro	880

[illegible]



Db 1546 ACCCTCTCACGACATCCAGCTGGCCACACACATGACATCCAGTCCGTCTG 1605  
 Qy 521 GLYSSERARGGLUGLARGPHEASPROSERALASERLEUASPHELEUTRPLACY 540  
 Db 1606 GGCAGACCCGGGAACAGAGGTTGACCCCTCTGCTCTGAGACTTCTTGCGGCGTGC 1665  
 Qy 541 ILEHISVALPROAQLILETRPGLNGLYARGASPGINARFTHPRGLNLSARGGLU 560  
 Db 1666 ATTCATGTTCTCGCATCTGGCAGGGGGGGGACACAGCCACCCCGACAGAGGGGGGAG 1725  
 Qy 561 GLULEUVALLEUARGVALGLNGLYPROGLULEUILESERLEUVALGLULEULEUAL 580  
 Db 1726 GAGCTGGTGTGCGGTCCAGGGCCCGGAGCTCATCAGCTGGTGGAGCTGATCTCTGCC 1785  
 Qy 581 GLUVALGLUPTHARGSERGLASPGLYSPHRAALACYSSELEULEGLINLARG 600  
 Db 1786 GAGGGGAGAGCGGAGGACGAGGACAGCCGCTGAGGCTCATCCAGGCCCGG 1845  
 Qy 601 LEUPROLEULEULEUSERCYSCYSGLYASPAJUSERVALARGLYVALTRGLU 620  
 Db 1846 CTGCCCCCTGCTAGCTGCTGCTGGGAGCATGAGAGTCAAGAGGTGACGGAG 1905  
 Qy 621 HISLEUSERGLYCYSLIENGLINTRPOLYASPSERVALLEUGLYARGARGCYSARGSP 640  
 Db 1906 CACCTGTAGGCTGCTATCCAGAGTGGGAGACACGCTGTGGAGGCGCTGCCGAGAC 1965  
 Qy 641 LEULEULEUGLNULEUTYRLEUGLNUARPPOGLULEARGVALPROVALPROGLUVALLEU 660  
 Db 1966 CTCTCCCTGACGCTCTACTCTACGAGCGGAGCTGGCGGTCGCCGTCTGAGGTCTCA 2025  
 Qy 661 LEUHSISERGLUGLYALALASERSESERVALCYSLYSLEUASPGLYLEULEHISARG 680  
 Db 2026 CTGCACACGGAAGGGGCTGCCAGCAGCAGCGTCTCAAGCTGGAGGACTCATCCAGCGC 2085  
 Qy 681 PHELEUTRLEULEULEALASPTHISERASPSERARGALALEUGLNUASARGLYALASP 700  
 Db 2086 TTCAATCAGCTCTCTTCCGAGACACACGACATCCCGGGCTTGGAGAACCGAGGGCGGAT 2145  
 Qy 701 ALASERMETALACYARGLYSLEUALAVALALAHISPROLEULEULEUARHISLEU 720  
 Db 2146 GCCAGCATGGCTGCGGAGCTGGCGGTGGCGACCCGCTCTCTGCTCAGGACCTG 2205  
 Qy 721 PROMETILEALALALEULEULEHISGLYARGTHRHSLEUASNPHEGLNGLUAPHEARGIN 740  
 Db 2206 CCCAATGATCGGGGCTCTCTGACGCGCGACCCACCTCAATCTCCAGAGACTTCCGGCAG 2265  
 Qy 741 GLINASNHSLEUSERCYSPHELEUHISVALLEUGLYLEULEUGLNULEUGLNUPROHS 760  
 Db 2266 CAGAACCACTGAGCTGCTCTCTGACGCTGGCGCTGTGAGCTCTCTGACGCCGAC 2325  
 Qy 761 VALPHEARGSERGLUHSISGLNGLYALALEUTRPAJPCYSLEULEUSERPHELEARGLEU 780  
 Db 2326 GTGTTCCCGCAGCAGCACAGGGGGGCTGTGGAGCTGCTCTCTCTCTCATCCGCTG 2385  
 Qy 781 LEULEUASNTRYARGLYSSESERARGHISLEUALAALAPHEILEASNLSPHEVALGIN 800  
 Db 2386 CTGCTGAATACAGSAGATCTCCCGCATCTGCTGCTTATCAACAAGTTTGACAG 2445  
 Qy 801 PHEILEHISLYSTYRILETHRTRYRASNALPROALALALILESERPHELEUCINLYSHS 820  
 Db 2446 TTCAATCCATTAAGTACATTACATCAATGCCCGACAGGCATCTCTCTGACAGAGAC 2505  
 Qy 821 ALAASPPROLEUHSISASPLEUSERPHEASPSANSERASPLEUVALMETLEUYSERLEU 840  
 Db 2506 GCCGACCGGCTCCACGACTGTCTTCGACAAACATGACTGTGTAATCCCTC 2565  
 Qy 841 LEUALAGLYLEUSERLEUPROSERARGASPARGTHASPARGLYLEUASPGIUGLU 860  
 Db 2566 CTTCGAGGGCTCAGCTCCACGAGGAGGACGACAGACCGAGGCTGACGACAGAG 2625  
 Qy 861 GLYGLUGLUGUSERSERALAGLYSERLEUPROLEUVALSERVALSERLEUPHERPRO 880  
 Db 2626 GCGGAGGAGAGAGCTCAGCGGCTCTTGGCCCTGTGCTGACGCTCTCTCTCTACCCCT 2685

Qy 881 LEUTHRALAAGLUMETALAPROTYRMETLYSARGLEUSERARGLYGLNTHRALGLU 900  
 Db 2686 CTGACCGGCGGAGATGGCCCCCTTACATGAACGGCTTTCCGGGGCCAAACGGTGGAG 2745  
 Qy 901 ASPLEULEUGLVALLEUSERASPILEASPGIUMETSERARGARGARGPROGLULEU 920  
 Db 2746 GATCTGTGAGAGTCTGTGATGACATPAGAGATGTCCCGCGGAGACCCAGATTCGTG 2805  
 Qy 921 SERPHEPSESERTHASNLEUGLINARGLEUMETSERVALAGLUGLYCYSARGASN 940  
 Db 2806 AGCTTCTTCTCGACCAACCTGTCAGCGGCTGATGACTCGGCGGAGAGTGTGGCGCAAC 2865  
 Qy 941 LEUALPHESEERLEUALLEUARGSERMETGINASNSERPROSERILEALALAPHE 960  
 Db 2866 CTGCGCTTACGCTGGCCCCCTGCTCTCATGACGAACAGCCCCACATTCACCGCTTTC 2925  
 Qy 961 LEUPROTHRPHMETTYRACYSLIENGLYSERGINASPPHEGLUVALGLINTHRALEU 980  
 Db 2926 CTGCCCCAGTTCATGACTGCTGGGACAGCAGACTTGAAGTGTGTGACAGACGGCCCTC 2985  
 Qy 981 ARGASNLEUPROGLUTYRALALEULEUCYSGINGLNUHISALALAVALLEULEHISARG 1000  
 Db 2986 CGGAACCTGCGAGTACGCTCTCTGTGCCAAAGACAGCAGCGCTGCTGCCACCGG 3045  
 Qy 1001 ALAPHELEUVALGLYMETTYRGLYGLINMETASPPROSERALAGLILESERGLUALLEU 1020  
 Db 3046 GCCTTCTGTGTGGGATGTACGAGGCCAGATGAGCCCGAGCGGCAGATCTCCAGGCCCTG 3105  
 Qy 1021 ARGILELEUHSMETGLUALAVALEMET 1029  
 Db 3106 AGCATCTGTGATATGAGAGCGCTGATG 3132  
 RESULT 4  
 ABK35138  
 ID ABK35138 standard; cDNA; 2824 BP.  
 AC ABK35138;  
 XX  
 DT 08-MAY-2002 (first entry)  
 DE Human cDNA encoding secreted protein #276.  
 KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;  
 KW viral infection; bacterial infection; fungal infection; diabetes; asthma;  
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;  
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;  
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;  
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;  
 KW tissue regeneration; wound healing; burn; haematopoiesis;  
 KW myeloid cell deficiency; lymphoid cell deficiency.  
 OS Homo sapiens.  
 PN WO200177288-A2.  
 PD 18-OCT-2001.  
 PE 29-MAR-2001; 2001WO-US10224.  
 PR 06-APR-2000; 2000US-195582P.  
 PA (GENY) GENETICS INST INC.  
 PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
 DR Gulukota K, Graham JR;  
 DR WPI: 2002-179321/23.  
 PT Five hundred and ninety two polynucleotides derived from a variety of  
 PT human tissue sources which encode secreted proteins, useful for  
 PT treating immune deficiencies and disorders such as autoimmune disorders

XX Claim 1, Page 210, 372pp: English.

PS The invention relates to 592 polynucleotides which have been derived from  
 CC a variety of human tissue sources and which encode novel secreted  
 CC proteins. The polynucleotides can be used as probes for the  
 CC identification and isolation of full length cDNA and genomic DNA. The  
 CC polynucleotides and proteins can also be used as nutritional supplements.  
 CC The proteins are useful in the treatment of various immune deficiencies  
 CC and disorders such as viral infections, bacterial infections, fungal  
 CC infections, autoimmune thyroiditis (e.g. rheumatoid arthritis, multiple  
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions  
 CC and conditions (e.g. asthma). They are also useful for treating  
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's  
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),  
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also  
 CC useful for tissue regeneration, for wound healing and in the treatment of  
 CC burns, incisions and ulcers. The proteins are also useful for regulating  
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.  
 CC Sequences ABK34853-ABK35454 represent polynucleotides of the invention.

XX Sequence 2824 BP; 443 A; 982 C; 910 G; 489 T; 0 other;

# Alignment Scores:

Pred. No.:	3,3e-302	Length:	2824
Score:	4211.00	Matches:	829
Percent similarity:	98.81%	Conservative:	1
Best Local Similarity:	98.69%	Mismatches:	0
Query Match:	80.35%	Indels:	10
DB:	24	Gaps:	1

US-09-929-769-7 (1-1029) x ABK35138 (1-2824)

QY 200 ProLeuSerProAspProArgTrpGlnSerSerProArgProValAlaLeuAlaLeu 219  
 DB 2 CGGCTCAGCCCGGAGCCCTGGGTGGAGAGCTCCAGTCCCGCCCGGCGCTCCGCTG 61  
 QY 220 GlnGlnAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 239  
 DB 62 CAGCAGAGCCCTGGGCGAGAGCTGGCCCGCTGCTCAGAGGAGCCCGGAGGCTCCGGG 121  
 QY 240 IleThrValArgValLeuGlnAlaLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 259  
 DB 122 ATACGGGTGGGTCTCTGCGAGGCTCCGACCTGCTGCTCCGACACGCGGAGTGGC 181  
 QY 260 LeuValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuCys 279  
 DB 182 CTGGTGTATGTCATGACCGCTAGCCACTTCTGCGCTGGCGCTGGCGCCAGCTCTGC 241  
 QY 280 GlnTyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValLeu 299  
 DB 242 CAGTACAGGCGTGTGTGGCAGACAGACCGGCTTCTCTCCGCTCTTCCGAGGCTCTC 301  
 QY 300 LeuGlnMetLeuGlnTrpLeuAspSerProGlyValGlnGlnGlnGlnGlnGlnGlnGln 319  
 DB 302 CTGCAGATGCTGCAGTGGCGGACAGCCCTGGGCTGGAGGCGGCGCCCTCCGCGGCA 361  
 QY 320 LeuArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGly 339  
 DB 362 CTCAGATGCTTCCAGCCAGGCGCTCAGCGGCGCGAGCTGATGTGTGCGAGGGGGG 421  
 QY 340 LeuLeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValValSerSerThr 359  
 DB 422 CTCTCTCGCTGGCGAGGCGCTGGCTTCCGTCAGACCTGGAGGCTGACCTCCACCC 481  
 QY 360 ValArgAlaValIleAlaThrLeuArgSerGlyGlnGlnCysSerValGlnProAspLeu 379  
 DB 482 GTCCGCGCCCTCATCCGACACCTGAGGTCTGGGGAGACAGTGCAGCTGGAGCCGAGCTG 541  
 QY 380 IleSerLysValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 399  
 DB 542 ATCAGCAAGTCTCTCAGGGGCTGATGAGGTAGAGTCCCGCCACCTGAGGAGGAGTCTCTG 601

QY 400 ThrAlaPheSerSerAlaThrAlaAspAlaSerProPheProAlaCysLysProVal 419  
 DB 602 ACTGATTTCTTCTCTGCGACCTGGGAGTGGCTCCCGCTTCCAGGCTGTAGCCCGGT 661  
 QY 420 ValValValSerSerLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 439  
 DB 662 GTGGTGGTGGTCTCTGCTGGCTGGCAGAGAGAGGCGCTGGGCTGGGGGAGAGCGGGT 721  
 QY 440 AlaAspGlyLysSerLeuGlnAlaValArgLeuGlyProSerSerGlyLeuLeuValAsp 459  
 DB 722 CGGAGAGGTGGCAGCCGAGAGGCGCTGGCGCTGGGGCGCTGTCAGGCTCTAGTGAGC 781  
 QY 460 TrpLeuGlnMetLeuAspProGlnValValSerSerCysProAspLeuGlnLeuArgLeu 479  
 DB 782 TGGCTGGAATGCTGGAGCCCGGAGGTGTGTCAGAGCTGCCACCGCAGCTCAGGCTG 841  
 QY 480 LeuPheSerArgTrpGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 499  
 DB 842 CTCTCTCCGAGAGAGAGGAG 901  
 QY 500 LeuThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeu 519  
 DB 902 CTGACCTCTTACAGCATCATGTCACAGCTGGCCACACTGCACAGTCCATCCGAGTCTG 961  
 QY 520 LeuGlyLysSerArgGlnGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAla 539  
 DB 962 CTGGGCAAGAGCCGGGAGACAGAGTTCGACCTCTCTCTGAGCTTCTCTGGGCC 1021  
 QY 540 CysIleHisValProArgIleTrpGlnGlyArgAspGlnArgTrpProGlnLysArgArg 559  
 DB 1022 TGCATTCATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081  
 QY 560 GluGlnLeuValLeuArgValGlnGlyProGlnLeuLeuSerLeuValGlnLeuLeu 579  
 DB 1082 GAGAGAGTGGTGTGGCGGCTGACAGGCGCGAGTCTACAGCTGGTGGAGCTGATCTG 1141  
 QY 580 AlaGlnAlaGlnThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAla 599  
 DB 1142 GCCGAGCGGAGAGCGCGGAGACAGAGGAGGAGACACACCCCTGACCTCATCCAGGCC 1201  
 QY 600 ArgLeuPheLeuLeuLeuSerCysCysGlyAspAspGlnSerValArgLysValThr 619  
 DB 1202 CGGCTGCCCTCTGCTCTCAACTGCTGCTGTGGGAGCATAGACTGTCAGAGAGTACG 1261  
 QY 620 GluHisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArgArgCysArg 639  
 DB 1262 GAGACACTGTGACGCTCATTCAGACAGACTGGGAGACAGCTGCTGGCGAGGCTCGCA 1321  
 QY 640 AspLeuLeuLeuGlnLeuTrpLeuGlnArgProGlnLeuArgValProValProGlnVal 659  
 DB 1322 GACCTTCTCTGACGCTCTACCTACACGCGCGAGCTGGGAGTGGCGCTGAGAGTC 1381  
 QY 660 LeuLeuHisSerGlnGlyAlaAlaSerSerSerValCysLysLysLeuAspGlyLeuHis 679  
 DB 1382 CTACTGCACAGCAAGAGGCTGCCAGCAGCAGAGGTCTGCAAGCTGAGGAGCATCATCC 1441  
 QY 680 ArgPheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAla 699  
 DB 1442 CGCTTCTACAGCTCTCTGGGAGACACAGCAGTCCCGGCGCTGGGAGAACCCAGGGGCG 1501  
 QY 700 AsnAlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuArgHis 719  
 DB 1502 GATGCCAGCATGGCTGCGGAGAGCTGGCGGTGGCGGAGCCGCTGGCTGCTAGGACAC 1561  
 QY 720 LeuProMetIleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPheGlnGlnPheArg 739  
 DB 1562 CTGCCATGATGCCGGGCTCTGTCAGCGGCGGAGCCACCTCACTTCCAGGAGTTCGG 1621  
 QY 740 GlnGlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuGlnPro 759  
 DB 1622 CAGCAGAACCACTGAGTCTCTGTCAGAGTCTGGGCTGAGGCTGAGCTGAGCAGCCG 1681  
 QY 760 HisValPheArgSerGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 779



```

|||||
Db 1682 CACGTGTTCCGACGAGCAGCAGGGGGCGCTGTGGGACCTCTTCTCTTCATCCGC 1741
Qy 780 LeuLeuLeuSerrArgArgSerSerArgHisLeuAlaIleAsnLysPheVal 799
Db 1742 CTGCTGTGAATTCAGGAAATCTCTCCGCGCATCTGCTGCTTCATCAACAGATTGTG 1801
Qy 800 GluPheIleHisLysTyrIleThrTyrAsnAlaProAlaIleSerPheLeuGlnLys 819
Db 1802 CAGTTCAATCCATAAGTACATTACTACATGCCAGAGGACCATCTCTTCTTCAGAGAG 1861
Qy 820 HisAlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSer 839
Db 1862 CACGCCGACCGCTCCACGACGCTCTTCCGACAAACAGTGCATGATGCTGCTGAATCC 1921
Qy 840 LeuLeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArgGlyLeuAspGlu 859
Db 1922 CTCTCTGCAAGGGCTCAGCTCTCCAGCAGGAGGAGCAGACCGAGGCTGTGACGAA 1981
Qy 860 GluGlyGluGluGluSerSerArgLysSerLeuProLeuValSerValSerLeuPheThr 879
Db 1982 GAGGGCCAGAGGAGAGAGCTACGCGGCTCTTCCCTGGGTCAGCGTCTCCCTGTTCAC 2041
Qy 880 ProLeuThrAlaAlaGluMetAlaProTyrMetLysArgLeuSerArgGlyGlnThrVal 899
Db 2042 CCTCTGACCGCGCGGAGATGGCCCTTACATGAACGGCTTTCGCCGGGCCAAGCGGTG 2101
Qy 900 Glu-----AspLeuLeuGluValLeuSerAspIle 909
Db 2102 GAGGGTGATGACGAGCCCTGCTTACCCAGCCAGATCTGTGGAGAGTCTAGTGACATA 2161
Qy 910 AspGluMetSerArgArgArgProGluIleLeuSerPhePheSerThrAsnLeuGlnArg 929
Db 2162 GAGCAGATATGTCGGGGGAGAGCCGAGATCCTGAGCTTCTCTGACCAACCTGAGGG 2221
Qy 930 LeuMetSerSerAlaGluGluCysArgAsnLeuAlaPheSerLeuAlaLeuArgSer 949
Db 2222 CTGATGAGCTCGCGGAGAGAGTGTGGCGCAACTCGGCTTCAGGCTGCGGCTGCC 2281
Qy 950 MetGlnAsnSerProSerIleAlaAlaIlePheLeuProIlePheMetTyrCysLeuGly 969
Db 2282 ATGCAGAACACCCCGAGATTCGACGCGCTTCTGCGCCACGCTCATCTACGCTGGGC 2341
Qy 970 SerGlnAspPheGluValIleGlnThrAlaLeuArgAsnLeuProGluTyrAlaLeuLeu 989
Db 2342 AGCCAGAGACTTGAAGTGTGTCCAGACGCGCCTCGGAGACGCTGATACGCTCTCTG 2401
Qy 990 CysGlnGluHisAlaAlaValLeuLeuHisArgAlaPheLeuValGlyMetTyrGlyGln 1009
Db 2402 TGCCAGAGACGCGGCGTGTCTGCTCCACCGGCGCTTCGTGGGCGATGACGCGCAG 2461
Qy 1010 MetAspProSerAlaGlnIleSerGlnAlaLeuArgIleLeuHisMetGlnAlaIleMet 1029
Db 2462 ATGGACCCGACGCGGAGATCTCCGAGGCGCTCGAGATCTGCAATATGAGGCGCTGATG 2521

RESULT 5
AAC76704
ID AAC76704 standard; cDNA; 2275 BP.
XX
AC AAC76704;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2259 polynucleotide sequence SEQ ID NO:4517.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnerary; antiparasitic; antiparkinsonian; neurotrophic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antiabiotic;
hypotensive; dermatological; immunosuppressive; antineuritic;
antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;

```

```

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
OS Homo sapiens.
XX
XX W0200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX
XX 02-APR-1999; 99US-0127636.
XX
XX 05-APR-1999; 99US-0127728.
XX
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkels RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX
XX P-PSDB; AAB42495.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 5; Page 3715-3716; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnerary;
XX antiparasitic; antiparkinsonian; neurotrophic; neuroprotective;
XX osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antiabiotic; hypotensive; dermatological; immunosuppressive;
XX antidiabetic; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 2275 BP; 393 A; 766 C; 719 G; 397 T; 0 other.
XX
XX
XX Alignment Scores:
XX Pred. No.: 1,43e-229 Length: 2275
XX Score: 3229.00 Matches: 633
XX Percent Similarity: 98.44% Conservative: 0
XX Best Local Similarity: 98.44% Mismatches: 0
XX Query Match: 61.61% Indels: 10
XX DB: 21 Gaps: 1
XX
XX US-09-929-769-7 (1-1029) x AAC76704 (1-2275)
XX
XX Qy 397 GluLeuLeuThrAlaPheSerSerArgThrAlaAspAlaIleSerProPheProLacys 416
XX |||||||
XX Db 22 GAGCTGCGACGACGATCTCTCTCTCCACTGCGGATGCTGCTCCCGCTTCCACCGTGT 81
XX |||||||
XX Qy 417 LysProValValValValSerSerLeuLeuGlnGluGluGluProLeuAlaGlyGly 436
XX |||||||

```

Db 82 AAGCCCTTGTGGTGGAGCTCCCTGCTGCTCAGAGAGAGAGCCCTGGCTGGGGG 141  
 QY 437 LysProGlyAlaAspGlyGlySerLeuGluAlaValArgLeuGlyProSerSerGlyLeu 456  
 Db 142 AAGCCGGGTGGAGCGGTGGAGAGCGGTGGGGCTGGGGCTGCTGAGGGCTC 201  
 QY 457 LeuValAspTrpLeuGlnMetLeuAspProGluValValSerSerCysProAspLeuGln 476  
 Db 202 CTAGTGGAGCTGGAAATGCTGAGCCCGGAGGTGGTCCAGCAGCTGCCCGACCTGAG 261  
 QY 477 LeuArgLeuLeuPheSerArgArgGlyGlyGluAlaGlnAlaGlnValProSerPheArg 496  
 Db 262 CTAGAGCTGCTCTTCTCCGAGAGAAAGGCAAGAGTCAAGGCCAGGGCTGGTCCGT 321  
 QY 497 ProTyrLeuLeuThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIle 516  
 Db 322 CCTACTCTCTGAGCCCTCTTACAGCATCAGTCCAGGTGGCCACAGTGCACAGTGTGATC 381  
 QY 517 ArgValLeuLeuGlyLysSerArgGluGlnArgPheAspProSerAlaSerLeuAspPhe 536  
 Db 382 CGAGTCTGCTGGGCAAGACCCGGGAACAGAGTTCCAGCCCTCTGCTCTGAGACTTC 441  
 QY 537 LeuTrpAlaCysIleHisValProArgIleTrpGlnGlyArgAspGlnArgTrpProGln 556  
 Db 442 CTCTGGGCTGCAATCCATGTTCTCCCATCTGGCAGGGGGGGAGCCACGCCCGGAG 501  
 QY 557 LysArgArgGluGluLeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGlu 576  
 Db 502 AAGCCGGGGAGAGCTGGTGGCTGGGGTCCAGGGCCCGAGCTCATACAGCTGGTGGAG 561  
 QY 577 LeuIleLeuAlaGluAlaGluThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeu 596  
 Db 562 CTATCTGGCCGAGAGCGGAGAGCGGAGAGCCAGAGCGGAGACACCCGCTCAGCCCTC 621  
 QY 597 IleGlnAlaArgPheProLeuLeuLeuSerCysCysGlyAspAspGlnSerValArg 616  
 Db 622 ATCCAGGCCCGGCTGCCCTGCTGCTCAGCTGCTGCTGGAGACCATGAGATGTCCAG 661  
 QY 617 LysValThrGlnHisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArg 636  
 Db 682 AAGGTACGAGAGCACTGTACAGGCTCATACACAGTGGGAGACAGCGTGGGCGAG 741  
 QY 637 ArgCysArgAspLeuLeuLeuGlnLeuTyrLeuGlnArgProGluLeuArgValProVal 656  
 Db 742 CGCTGCCGAGACTTCTCTGCACTGCTTACTACAGCGGCGAGCTGGGGGCCCGT 801  
 QY 657 ProGluValLeuLeuHisSerGlnGlyAlaAlaSerSerValCysLysLeuAspGly 676  
 Db 802 CCGAGGCTCTACTGCACAGCGAAGGGGCTGCCAGCAGCGCTGCCAAGCTGGAGCA 861  
 QY 677 LeuIleHisArgPheIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnAsn 696  
 Db 862 CTCTATCCACCGCTTCTATCAGGCTCTTGGGAGACACAGCACTCCGGGGCTGGAGAC 921  
 QY 697 ArgGlyAlaAspAlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeu 716  
 Db 922 CGAGGGGGGAGATCCAGCATGAGCTGCCGAGACTGGCGGTGGCGACCCGCTGCTGCTG 961  
 QY 717 LeuArgHisLeuProMetIleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPheGln 736  
 Db 982 CTGAGGCACTGCCCATGATCGGGGCTCTCTGACGGCGGACCCACGTCATTCCAG 1041  
 QY 737 GluPheArgGlnGlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGluLeu 756  
 Db 1042 GAGTTCTGGGAGAGCAACCACTGAGCTCTCTCTGAGTGGCTGGCTGGAGAGCTG 1101  
 QY 757 LeuGlnProHisValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSer 776  
 Db 1102 CTGACAGCCCACTGTTCCGAGCGAGCAACAGGGGGCTGGGAGCTGCTTCTGCTC 1161  
 QY 777 PheIleArgLeuLeuLeuAsnTrpArgLysSerSerArgHisLeuAlaPheIleAsn 796  
 Db 1162 TACATCGGCTGCTGTAATTACAGAGAGTCTCTCCGCACTGCTGCTCATCAAC 1221

QY 797 LysPheValGlnPheIleHisIleTyrIleThrTyrAsnAlaProAlaAlaIleSerPhe 816  
 Db 1222 AAGTTTGTGCAATTTCATTCATAGTATCATCTACATGCCCCAGAGCATCTCTTC 1281  
 QY 817 LeuGlnLysHisAlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMet 836  
 Db 1282 CTGCAGAGACCGCCCGCCGCTCCAGACCTTCTCTTCCACAAAGTGCACCTGGTATG 1341.  
 QY 837 LeuLysSerLeuLeuAlaGlyLeuSerLeuProSerArgAspAspArgTrpAspArgGly 856  
 Db 1342 CTGAATTCCTCTTGCAGGGCTCAGCTGAGCCAGAGGAGACAGAGCCAGAGG 1401  
 QY 857 LeuAspGluGluGlyGlnGluGlnGlnSerSerAlaGlySerLeuProLeuValSerValSer 876  
 Db 1402 CTGAGAGAAAGAGGCGAGAGAGAGAGTCCAGCCGGCTCTGGCCCTGGTCCAGCTTC 1461  
 QY 877 LeuPheThrProLeuThrAlaAlaGlnMetAlaProTyrMetLysArgLeuSerArgGly 896  
 Db 1462 CTGTTCACTCCCTTGACCGCGCGGAGATGGCCCTTACATGAAGAGGCTTCCCGGG 1521  
 QY 897 GlnThrValGlu-----AspLeuLeuGluValLeu 906  
 Db 1522 CAACGGTGGAGGGTGAAGTACAGCCCTGCTTACACCCAGCCAGATGCTGAGAGTTCTG 1581  
 QY 907 SerAspIleAspGlnMetSerArgArgArgProGluIleLeuSerPhePheSerTrpAsn 926  
 Db 1582 AGTACATAGACAGATGATGTCCTCCGCGAGACCCGAGATCTCTGAGCTTCTCGACCAAC 1641  
 QY 927 LeuGlnArgLeuMetSerSerAlaGluGluCysCysArgAsnLeuAlaPheSerLeuAla 946  
 Db 1642 CTCCAGGGCTGTAGAGCTGCGCGAGAGGTGTGGCGCAACTCCGCTTCCAGCTGGCC 1701  
 QY 947 LeuArgSerMetGlnAsnSerProSerIleAlaAlaAlaPheLeuProThrPheMetTyr 966  
 Db 1702 CTGCGCTTCATGAGAAAGCCCGACCATTTGACGGCTTCTGCGCACGTTCAATGAC 1761  
 QY 967 CysLeuGlySerGlnAspPheGluValGlnThrAlaLeuArgAsnLeuProGluTyr 986  
 Db 1762 TGCCTGGGCGCCAGGACTTTGAGTGTGTCACAGCGCCCTCCGGAACCTGCTGAGTAC 1821  
 QY 987 AlaLeuLeuCysGlnGlnHisAlaAlaValLeuLeuHisArgAlaPheLeuValGlyMet 1006  
 Db 1822 GCTCTCTGTGCCAAGAGACCGCGGCTGTGCTCTCCACCGGCTTCTGCTGGGCAAG 1881  
 QY 1007 TyrGlyGlnMetAspProSerAlaGlnIleSerGlnAlaLeuArgIleLeuHisMetGln 1026  
 Db 1882 TACGGCCAGATGGACCCAGCGCGAGATCTCCGAGGCGCTGAGATCCGATATGAG 1941  
 QY 1027 AlaValMet 1029  
 Db 1942 GCCGTGATG 1950  
 RESULT 6  
 AA298062  
 ID AA298062 standard; cDNA; 1559 BP.  
 AC AA298062;  
 XX  
 DT 09-MAY-2000 (first entry)  
 XX  
 DE Human secreted protein encoding nucleotide sequence SFD ID NO:56.  
 XX  
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
 KW antiinflammatory; nocotropic; neuroprotective; antiallergic; cancer;  
 KW tumour; neurodegenerative disorder; developmental abnormality; allergy;  
 KW foetal deficiency; blood disorder; immune system disorder; arthritis;  
 KW autoimmune disease; hepatic disease; renal disease; inflammation;  
 KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;  
 KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;  
 KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;  
 KW reproductive disorder; gastrointestinal disorder; respiratory disorder;  
 KW metabolic disorder; food additive; preservative; ss.





DB 1201 GAGCCGTCATG 1212  
RESULT 7  
AAD11675  
ID AAD11675 standard; cDNA; 1559 BP.  
XX  
XX AAD11675;  
XX  
XX 24-SEP-2001 (first entry)  
DE Human secreted protein-encoding gene 46 cDNA clone HBCY35, SEQ ID NO:56.  
XX  
XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;  
KM foetal abnormality; developmental abnormality; haematopoietic disorder;  
KM immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KM parkinson's disease; cognitive disorder; schizophrenia; skin disorder;  
KM psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;  
KM inflammation; neurological disorder; Alzheimer's disease; food additive;  
KM angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;  
KM pregnancy-related disorder; endocrine disorder; infection; wound healing;  
KM cell culture; chemotaxis; vulnery; binding partner identification;  
KM gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 232..1215  
FT /tag= a  
FT /product= "Human secreted protein precursor"  
FT sig\_peptide 232..300  
FT /tag= b  
FT mat\_peptide 301..1212  
FT /tag= c  
FT /product= "Mature human secreted protein"  
XX  
PN WO200151504-A1.  
XX  
XX 19-JUL-2001.  
XX  
XX 12-JAN-2001; 2001WO-US00911.  
XX  
XX 13-JAN-2000; 2000US-0482273.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y,  
PI Lafleur DM, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;  
PI Address GA, Muscenski M, Ebner R;  
XX  
XX WPI: 2001-425865/45.  
DR P-PSDB: AAE06086.  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is  
PT used in preventing, treating or ameliorating a medical condition  
XX  
XX Claim 1; Page 704; 864pp; English.  
XX  
XX AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted  
CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.  
CC AAE06133-AAE06205 represent human secreted protein fragments.  
CC The secreted proteins and their genes are useful for preventing, treating  
CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
CC Pathological conditions can be diagnosed by determining the amount of the  
CC new protein in a sample or by determining the presence of mutations in  
CC the new genes. Specific uses are described for each of the 71 genes,  
CC based on the tissues in which they are most highly expressed, and include  
CC developing products for the diagnosis or treatment of proliferative  
CC disorders, cancer, tumours, foetal and developmental abnormalities,  
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,

CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
CC pregnancy-related disorders, endocrine disorders, and infections. The  
CC proteins can also be used to aid wound healing and epithelial cell  
CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
CC before transplantation, for supporting cell culture of primary tissues,  
CC to regenerate tissues, to identify their cognate ligands or binding  
CC partners, and in chemotaxis, and can be used as a food additive or  
CC preservative to modify storage properties. Antibodies specific for a  
CC protein of the invention can be used in alleviating symptoms associated  
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).  
CC The present sequence represents a human secreted protein-encoding cDNA of  
CC  
XX  
XX  
SQ Sequence 1559 BP; 281 A; 530 C; 478 G; 267 T; 3 other:  
XX  
XX  
Alignment Scores:  
Pred. No.: 8,36e-143 Length: 1559  
Score: 2055.00 Matches: 404  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 39.21% Indels: 0  
DB: 22 Gaps: 0  
US-09-929-769-7 (1-1029) x AAD11675 (1-1559)  
QY 626 IIEGInGInTPGlyAspSerValLeuGlyArgArgCysArgAspLeuLeuGInLeu 645  
DB 1 ATCCACAGAGTGGGAGACAGCGTGTGGGACGGCTGGCCGAGACTCTCTCGAGCTC 60  
QY 646 TTTLeuGInArgProGluLeuArgValProValProGluValLeuLeuHisSerGluGly 665  
DB 61 TACCTACAGCGGCGGAGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGG 120  
QY 666 AlaAlaSerSerSerValCysLysLysAspGlyLeuIleHisArgPheIleThrLeuLeu 685  
DB 121 GCTGCCAGACAGAGCTCTGCAAGCTGAGACGAGACTATCCACCGCTTCAACGCTCTT 180  
QY 686 AlaAspThrSerAspSerArgAlaLeuGluAsnArgGlyAlaAspAlaSerMetAlaCys 705  
DB 181 GCGACACAGACGACACGCGCGGTGGAGAACCGAGGCGGATGCCAGCATGGCTCGC 240  
QY 706 ArgGlyLeuAlaValAlaHisProLeuLeuLeuLeuArgHisLeuProMetIleAlaIle 725  
DB 241 CGGAAGCTGGCGGTGGCGACCGCTGCTGCTGAGCACCTGCCATGATGGCGGCG 300  
QY 726 LeuLeuHisGlyArgThrHisLeuAsnProGInGInPheArgGInGInAsnHisLeuSer 745  
DB 301 CTGCTGACAGGCGGCGACCCACCTCACTCCAGATTCGCGAGCAGAACCCAGCTGAGC 360  
QY 746 CysPheLeuHisValLeuGluLeuLeuGluLeuLeuGInProHisValPheArgSerGlu 765  
DB 361 TGCCTTCCTGACAGCTGCTGGGCGCTGCTGGAGCTGCTCACCGCAGCTGTCCGACGAG 420  
QY 766 HisGInGlyAlaLeuThrAspCysLeuLeuSerPheIleArgLeuLeuAsnTyrArg 785  
DB 421 CACGAGGAGGCGCTGGGAGCTGCTGTGCTTATCGGCTGCTGATTTACAGG 480  
QY 786 LysSerSerArgHisLeuAlaPheIleAsnLysPheValGInPheIleHisLysTyr 805  
DB 481 AAGCTCTCCCGCATGTGGCTGCTTCATCAACAACTTTGTGAGTTCATTAAGTAC 540  
QY 806 IleThrTyrAsnAlaProAlaAlaIleSerPheLeuGInLysHisAlaAspProLeuHis 825  
DB 541 ATTACCTCAATAGCCCCACAGCATCTCTCTCCACAGACACCGCGCTCCAC 600  
QY 826 AspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeuSer 845  
DB 601 GACCTGCTCTTGACAAACAGTACCGTGAAGTGAATCCCTCTCCAGGCTCCAC 660  
QY 846 LeuProSerArgAspAspArgThrHisPArgGlyLeuAsnGlyGInGlyGInGInLys 865  
DB 661 CTGCCAGCAGGAGCAGACGAGCCGAGCGCTGAGCAGAAAGAGGCGAGAGGAGAC 720

```
QY 866 SerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThrAlaAlaGlu 885
DB 721 TCACCGCGGCTCTGCGCCCTGCTGAGCGCTCCCTGTTCCACCCCTCTACCGCGCGAG 780
QY 886 MetAlaProTyrMetLysArgLeuSerArgGlyInThrValGluAspLeuGluVal 905
DB 781 ATGGCCCCCTACATGAAGACGGCTTCCCGGCGCAACGGCTGGAGATCTCTCGAGGTT 840
QY 906 LeuSerAspLeuAspGluMetSerArgArgArgProGluIleLeuSerPhePheSerThr 925
DB 841 CTGAGTACACATGACGAGATGTCCCGGCGAGACCCGAGATCTGAGCTTCTTCGACG 900
QY 926 AsnLeuGlnArgLeuMetSerSerSerAlaGluGluCysCysArgAsnLeuAlaPheSerLeu 945
DB 901 AACCTGACGGCGCTGATGAGCTCGCGCGAGAGTGTTCGCCCACTCGCTTACGCTG 960
QY 946 AlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaPheLeuProThrPheMet 965
DB 961 GCCCTGGCGCTCCATGACGAGAACGCCCGACATGTCAGCCCTTCTTCCACGTTCAATG 1020
QY 966 TyrTyrLeuGlySerGlnAspPheGluValValGlnThrAlaLeuArgAsnLeuProGlu 985
DB 1021 TACTGCTGCGGACGACGAGCTTGTGAGGTGTGACAGAGCGCCCTCGGACCTGCTGAG 1080
QY 986 TyrAlaLeuLeuGlyGlnGluAlaAlaValLeuLeuHisArgAlaPheLeuValGly 1005
DB 1081 TACGCTCTCCGCTGCCAAGACACGCGGCTGTGCTGCTCCACCGGCTTCTGCTGAG 1140
QY 1006 MetTyrGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeuArgIleLeuHisMet 1025
DB 1141 ATGATCGCGCCAGATGAGACCCGACGCGCAGATCTCCGAGCGCTGAGATCTCGATATG 1200
QY 1026 GluAlaValMet 1029
DB 1201 GAGGCCGTGATG 1212

RESULT 8
ABK69771
ID ABK69771 standard; cDNA; 1559 BP.
AC ABK69771;
XX 15-JUL-2002 (first entry)
DE Human secreted protein gene 46.
XX
KW Human; ss; gene; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW vitruclide; fungicide; ophthalmological; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
XX epithelial cell proliferation; food additive.
XX Homo sapiens.
XX OS
XX NO200226931-A2.
XX
XX PD 04-APR-2002.
XX
XX PF 24-SEP-2001: 2001WO-US29871.
XX
XX PR 25-SEP-2000: 2000US-234925P.
XX
XX PR 12-JAN-2001: 2001WO-US00911.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y,
XX Lafleur DM, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR,
XX Endress GA, Mucenski M, Ebner R;
```

```
XX WPI: 2002-362489/39.
DR P-PSDB: ABG33908, ABG33997.
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma
XX
XX Claim 1: Page 1187-1188: 1478pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (or its
XX fragment, homologue complement or allelic variant) encoding a human
XX secreted protein (and its fragment, domain, epitope, variant, secreted
XX form and species variant). Also included are a recombinant vector
XX comprising the nucleic acid, a recombinant host cell comprising the
XX vector, an antibody against the secreted protein, a recombinant host cell
XX that expresses the secreted protein and a method of identifying a binding
XX partner of the secreted protein. The nucleic acid and protein are used to
XX prevent, diagnose, treat or ameliorate a medical condition in e.g.
XX humans, mice, rabbits, goats, cats, dogs, chickens or sheep
XX for example autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi and
XX ocular disorders e.g. corneal infection. Many other diseases and
XX disorders are listed in the specification. The polypeptides can also be
XX used to aid wound healing, an epithelial cell proliferation, to prevent
XX skin aging due to sunburn, to maintain organs before transplantation, for
XX supporting cell culture of primary tissues, to regenerate tissues and in
XX chemotaxis. The polypeptides can also be used as a food additive or
XX preservative to increase or decrease storage capabilities. The present
XX sequence encodes a novel human secreted protein of the invention.
XX
XX S0 Sequence 1559 BP; 281 A; 530 C; 478 G; 267 T; 3 other:
XX
XX Alignment Scores:
XX Pred. No.: 8.36e-143 Length: 1559
XX Score: 2055.00 Matches: 404
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 39.21% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-929-769-7 (1-1029) x ABK69771 (1-1559)
QY 626 IleGlnGlnTyrGlyAspSerValLeuGluGlyArgArgCysArgAspLeuLeuGlnLeu 645
DB 1 ATCCAGCAAGTGGGAGACAGCTGCTGGGACGCGCTGCCGAGACTTCTCTGCAAGTC 60
QY 646 TyrLeuGlnArgProGluLeuArgValProValProGluValLeuLeuHisSerGluGly 665
DB 61 TACCTACAGCGGCGGAGCTGGGCTGGCGCTGAGAGTCTCTAGCAGCAAGAGGG 120
QY 666 AlaAlaSerSerValCysLysLeuAspGlyLeuIleHisArgPheIleThrLeuLeu 685
DB 121 GCTGCCACAGAGCGCTGCAAGCTGGAGCGACATCATCCAGGCTATCATCGCTCTT 180
QY 686 AlaAspThrSerAspSerArgAlaLeuGluAsnArgGlyAlaAspAlaSerMetAlaCys 705
DB 181 GCGGACACGAGGAGCTCCCGGCGCTTGGAGAACGAGGCGCGATGCCAGCATGGCTCC 240
QY 706 ArgLysLeuAlaValAlaHisProLeuLeuLeuLeuArgHisLeuProMetIleAlaIle 725
DB 241 CGGAGAGCTGGCGGCGCGACCCGCTGCTGCTGCTGAGGACCTGCCATATGCGCGGCG 300
QY 726 LeuLeuHisGlyArgThrHisLeuAsnPheGlnGluPheArgGlnGlnAsnHisLeuSer 745
DB 301 CTCCTGACAGGCGCGACCCACTCAATCTTCAGGAGTTCCGCGACAGAACCACTGAGG 360
QY 746 CysPheLeuHisValLeuGlyLeuLeuGluLeuLeuGlnProHisValPheArgSerGlu 765
DB 765
```





OY 177 ArgValGlyThrGlnLeuArgValLeuGlyProGlnAsp 190  
 DB 44 CCGGTGGGACCCAGCTCCGGGTGGGCCCTGAGGACGAC 3  
 RESULT 11  
 ID ABL05745 standard; cDNA; 6207 BP.  
 XX ABL05745;  
 AC ABL05745;  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11717.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001: 2001WO-US09231.  
 XX  
 PR 23-MAR-2000: 2000US-191637P.  
 PR 11-JUL-2000: 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 XX  
 DR MPI: 2001-656860/75.  
 DR P-PSDB: ABB61642.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PT  
 PS Claim 1: SEQ ID NO 11717; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB161617), expressed DNA  
 CC sequences (AB101840-AB161617) and the encoded proteins  
 CC (ABBS7737-ABBS7202).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/publised\_pct\_sequences.  
 XX  
 SO Sequence 6207 BP; 1572 A; 1715 C; 1543 G; 1377 T; 0 other:  
 Alignment Scores:  
 Pred. No.: 1.2e-53 Length: 6207  
 Score: 863.00 Matches: 288  
 Percent Similarity: 43.40% Conservative: 189  
 Best Local Similarity: 26.21% Mismatches: 400  
 Query Match: 16.47% Indels: 222  
 DB: 23 Gaps: 34  
 US-09-929-769-7 (1-1029) x ABL05745 (1-6207)  
 OY 1 MethHisIleuValValHisAlaMetValIleuLeuThrLeuGlyProProArgAla 20  
 DB 3328 CTCACATTAACATCATTCACGCTTCATATCTGCTACCTATTC-----AACAGC 3381  
 OY 21 AspAspSerGluPheGlnAlaLeuLeuAspIleTirPhePheGluGluValProLeuPro 40  
 DB 3382 AACATGCCGAATCGATTACCATTTTGACTATTGGTTCCGCCGAGACGACGAGCC 3441

OY 41 ThrAlaPheLeuValAspThrSerGluGlnAlaLeu---LeuLeuProAspTirPleuLys 59  
 DB 3442 GTGCGCTTCTGCTCCAGCATCGCCGAGACGAGAGTGCACATCTGCTCCGCTGAGG 3501  
 OY 60 LeuArgMetIleArgSerGluValLeuArgLeuValAspAlaIleuGlnAspLeuGlu 79  
 DB 3502 CTGAAGATGATCCGCTATCGGACAGCTATGAGGACGCTCTCAACAGCTCAACG 3561  
 OY 80 ProGlnIleuLeuLeuPheValGlnSerPheGlyIleProValSerMetSerLys 99  
 DB 3562 CCGGATCAGATCGTGTCTTGTGCAAACTTGGCAGCGCGGTCTCAACTGATGCTCAAG 3621  
 OY 100 LeuLeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGlnAsnIle 119  
 DB 3622 CTGCTGCAATGCTGGACACCGCTGACTGAGACAGTTGATGCTGGTGAAGATGCCATT 3681  
 OY 120 MetAspLysAsnTyrMetAlaHisIleValGlnValGlnHisGluArgGlyAlaSerGly 139  
 DB 3682 CTGAACAGGCTATCTGGCCAGCTATGCAATTCAGACGCGCGGTCTTAAGAT 3741  
 OY 140 GlyGln---ThrPheHisSerLeuLeuThrAlaSerLeuProProArgAspSerThr 158  
 DB 3742 GGGCACTACACCGTACAGCGCTTGATCTGCATTC----- 3777  
 OY 159 GluAlaProLysProLysSerSerProGlnIleProIleGlyGlnGlyArgIleArgVal 178  
 DB 3778 -----CACTGCAGACTGTGCAGACTGCC-----AAGATCACTGTG 3816  
 OY 179 GlyThrGlnLeuArgVal-----LeuGlyProGluAspLeuAla 192  
 DB 3817 GTTATTTCAGGAGCGCGTTGAATGATGATTACGATTCTTCAGACTGCGGATGACGCC 3876  
 OY 193 GlyMetPheLeu-----GlnIlePheProLeuSerProAspProArg 206  
 DB 3877 ACTAATCTCTGCGCCACCAAGAGTAGCCCAACCATCTCAGCAGCCGAC----- 3930  
 OY 207 TyrGlnSerSerSerProArgProValAlaLeuAlaLeuGlnGlnAlaLeuGlyGlnGlu 226  
 DB 3931 ---CAGTTGACTGAGTGCAGAGTGCAGTCCCTTCATTCAAAACAGTGTGGACATG 3987  
 OY 227 LeuAlaArgValValGlnGlySerProGlnValProGlyIleThrValArgValLeuGln 246  
 DB 3988 CTAGCAAGCCCGAAGCAATAGAGCGGATGTGTACGCCCAACAGAGGTGTACGCC 4047  
 OY 247 AlaLeuAlaThrLeuLeuSerSerProHisGlyGlyAlaLeuValMetSerMetHisArg 266  
 DB 4048 GTG-----GTTGCAAGTGTCAAGATGAGCCGTAC----- 4077  
 OY 267 SerHisPheLeuAlaCysProLeuLeuArgGlnLeuCysGlnTyrGlnArgCysValPro 286  
 DB 4078 -----GCCGTGCACATTTTAAAG----- 4095  
 OY 287 GlnAspThrGlyPheSerSerLeuPheLeuValLeuLeuGlnMetLeuGlnTyrPhe 306  
 DB 4096 -----ACTTTCTTCAGCTGCATCTGCACAGCGACAGATACATCTGGAG----- 4143  
 OY 307 AspSerProGlyValGlnGlyGlyProLeuArgAlaGlnLeuArgMetLeuAlaSerGln 326  
 DB 4144 -----AACGCTCTCCAAAGAACTTAGTATGTT----- 4173  
 OY 327 AlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeuLeuArgLeuAlaGlnAla 346  
 DB 4174 -----AAGCACACATTCGCGGACCTCCAGCTCTGCTGCAGAAATCCGAA----- 4215  
 OY 347 LeuAlaPheArgGlnAspLeuGluValValIleSerSerThrValArgAlaValIleAlaThr 366  
 DB 4216 ---CTTATCATAGAGACTTGGTTCATCTGTGAGAAACTCTGCGAGATTATGTGG--- 4269  
 OY 367 LeuArgSerGlyGlnGlnCysSerValGluProAspLeuIleSer-----LysValLeu 384  
 DB 4270 -----CAGCAGTTTAAGGGGAAATACCGCTTGTGTGCGACGAAACGAGATCGTC 4317

QY 385 GlnGlyLeuLeuLeuValArgSerProHisLeuGluLeuLeuThrAlaPhePheSer 404  
 Db 4318 CGGGCGATCGTCCAA-----AGCTTCGAT 4341  
 QY 405 AlaThrAlaAspAlaAlaSerProPheProAlaCysLysProValValValSerSer 424  
 Db 4342 CAGACCAAGGAGAGC-----AAGACCGTCCCAAGTCCAGAC 4380  
 QY 425 LeuLeuLeuGlnGluGluProLeuAlaGlyGlyLysProGlyAlaAspGlyLysSer 444  
 Db 4381 GACGACGCTTCCAC----- 4395  
 QY 445 LeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuGluLeu 464  
 Db 4396 -----AACGGGCTCTTCATCGACTGCTGCTCCGAATG 4428  
 QY 465 AspProGluValValSerSerCysProAspLeuGluLeu-----ArgLeuLeuPhe 481  
 Db 4429 GATCCCGAGATAGTTTCCCT-----CAGCTAATGAAAGAGCGCTTCTGTTC 4476  
 QY 482 SerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProTrpLeuLeuThr 501  
 Db 4477 TCAGAG-----TCTCGCAGGAGTTTAGTTCATCTGTTC 4515  
 QY 502 LeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysLysLeuArgValLeuLeuGly 521  
 Db 4516 CTGATTCACCAACCAACCACTGGGACGACATCGAAGG-----ATTCCGAGTACCTGTTC 4572  
 QY 522 LysSerArgGluGlnArgPheAspProSerAlaSerLeuAspPheLeuThrAlaCysLys 541  
 Db 4573 AAGAAATTCATGAAGACTACGACTACGCCCGTCTCACTCACTTCGAGGACTGAC 4632  
 QY 542 HisValProArgLysLeuGlnGlnArgAspGlnArgThrProGlnLysArgArgGluGln 561  
 Db 4633 ACCAATCCGAGCTGCGAAGGAGCGGACGACATCATTCGAAGAAGCTACGCCGAT 4692  
 QY 562 LeuValLeuArgValGlnGlyProGluLeuLysLeuValGluLeuLeu----- 578  
 Db 4693 GCGTCTTCATCTGAGGACTTCTGAACTGGAGCCGTTCTCCACTTCACCTTCACGAA 4752  
 QY 579 -----LeuAlaGluAlaGlnThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeu 597  
 Db 4753 GGGCTTTCGAGGCTCAAGTGCACGCAAAACTATGATTTCAAGCTGTGCTG----- 4806  
 QY 598 GlnAlaArgLeuProLeuLeuLeuSerCysCysGlyAspAspLysLeuValArgLys 617  
 Db 4807 -----CGAATGACCTACTGTTTAAGTCACAGAGAACGACGACCTCATGTGCAAG 4860  
 QY 618 ValThrGlnHisLeuSerGlyCysLysGlnGlnThrProGlyAspSerValLeuGlyArg 637  
 Db 4861 GTAATGGACCGACGTGAG-----AAAAGTTCGGTTCGCACTACTTG 4902  
 QY 638 CysArgAspLeuLeuLeuGlnLeuTrpLeuGlnArgProGluLeuArgValProValPro 657  
 Db 4903 AAACATCGAGGCTCTCGACGATGATCATGTATCCGCGCATGAAGTTTCTGAACCG 4962  
 QY 658 -----GluValLeuLeuHisSerGlnGlyAlaAlaAspSerSer 670  
 Db 4963 GCGAAACCGGCGAGCGGCTGACAAATTCGCAATCTGAAGGC----- 5007  
 QY 671 ValLysLysLeuAspCysLysLeuLeuHisArgPheLeuLeuAlaAspThrSerAsp 690  
 Db 5008 -----TGCAGCGCGAGCAAGGTTCGCAAACTGATCCTGTGGCGAGTGTGGGCG 5064  
 QY 691 SerArgAlaLeuGlnLysAsnArgCysAlaAspAlaSerMetAlaCysArgLysLeuAlaVal 710  
 Db 5065 AAAAAGACTTTGAAACCTGTTCACGAGGACCGAGCTGCTTTCGCAAGCTGGTGGC 5124  
 QY 711 AlaHisProLeuLeuLeuLeuArgHisLeuProMetLysAlaAlaLeuLeuHisLysArg 730  
 Db 5125 TCCACACCGCTCTCTCTGCGCAGCTGGGCGTGTGCGTCATTTATGCAAGTGG 5184  
 QY 731 ThrHisLeuAsnPheGlnGluPheArgGlnGlnAsnHisLeuSerCysPheLeuHisVal 750

Db 5185 GCCCAGCTGAGCATGAAGCCCTGCCGAGAGACACACTTCCACCGATTTGTACAGATC 5244  
 QY 751 LeuGlyLeuLeuGlnLeuLeuGlnProHisValPhe-----ArgSerGlnHis 766  
 Db 5245 CTGAGGAGCGCTGAGTGTGTGCAACCACTTTTGGAGAGGCGTACGAAGCGATGC 5304  
 QY 767 GlnGlyAlaLeuThrAspCysLeuLeuSerPheLeuArgLeuLeuLeuAsnTrpArgLys 786  
 Db 5305 CAAAACAGCTG-----TCGTGCTACTTCACACTTCTTCAAG-----CACCAAC 5349  
 QY 787 SerSerArgHisLeuAlaAlaPheLeuAsnLysPheValGlnPheLeuHisLysTrpLeu 806  
 Db 5350 AACGTAAAGAGCGCTGCAAAATGCTGAAACAACTTTGTGCAGATGCTCCAAAGCTACATC 5409  
 QY 807 ThrTrpAsnAlaProAlaAlaLeuSerPheLeuGlnLysHisAlaProLeuHisAsp 826  
 Db 5410 AACTCAAAATCCCTCAAGTCTGTCTTCTTCATGCAAGACTGAGCTGGCATCTCGAAGAA 5469  
 QY 827 LeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeuSerLeu 846  
 Db 5470 CTTCGACCAAGTACACCTCCTCAGGCAAACTCAGATTTGTGTCAGGCCGTGCTG 5529  
 QY 847 ProSerArgAspArgThrAspArgGlyLeuAspGluGly----- 861  
 Db 5530 CTGCACACAAAGTCGACCTCGCGCAGCAAGTATGACAGAGAAAGTAAAGTACAGATAC 5589  
 QY 861 ----- 861  
 Db 5590 GATCTGGATGAGCATTTGATGATGAAGCCATGCGCCAGAACCCCGTGTACAGAGAT 5649  
 QY 862 -----GluGluGlnSerSerAlaGlySerLeuPro 871  
 Db 5650 CCCATGAGTGAATCCGAACACCACTCATGATCCAGCAGCATGAGGGTCCCTATG 5709  
 QY 872 LeuValSerValSerLeuPheThrProLeuThrAlaAlaGluMetAlaProTrp----- 889  
 Db 5710 GTTCTAACCTTGGGCTGCTGACACCGCGTCACTACAGGACATATGCGCGCTCTCTC 5769  
 QY 890 -----MetLysArgLeuSerArgGlyGlnThrValGlnAspLeuGluValLeuSer 907  
 Db 5770 GATCTGTCAGATCATTAAGCAAGTCCAAACGAGACGAGTGTGCTGGTCCATGAC 5829  
 QY 908 AspLeuAspLysMetSerArgArgArgProGluLeuLeuSerPhePheSerThrAsnLeu 927  
 Db 5830 GAGCTGAGTGGCTCACTTCCAAAGATTTGTGTCATCAACAGAGCTGTTCGAACGATG 5889  
 QY 928 GlnArgLeuMetSerSerAlaGluGlyCysArgAsnLeuAlaPheSerLeuAlaLeu 947  
 Db 5890 CTCACCTTATATTCGCGCGAGTGCAGATCCGCTCCATGCTCATCATCATGATC 5949  
 QY 948 ArgSerMetGlnAsnSerPro-----SerLysAlaAlaAlaPheLeuProThrPhe 964  
 Db 5950 AGGCATCTGAAGACATCCCGCAATTCGGACATCAACCTGTGCACTTAAAGCTTAC 6009  
 QY 965 MetTrpCysLeuGlySerGlnAspPheGluValGlnThrAlaLeuArgAsnLeuPro 984  
 Db 6010 ATTCACTGCTGGGAGGACGACCTCGTGGTGGAGCGACGACCATTAACATGATGCGG 6069  
 QY 985 GluTrpAlaLeuLeuGlnGlnGlnHisAlaAlaValLeuLeuHisArgAlaPheLeuVal 1004  
 Db 6070 GAGATGCGGTGCTGTGGAGGAAACGCAATGCAATCCTTAACGGTGTGCTGCTG 6129  
 QY 1005 GlyMetTrpGlyGlnMetAspProSerAlaGlnLysSerGlnAlaLeuArgLysLeu 1023  
 Db 6130 GCGTTGAAGTCGTGCTGAACACTGGCCACGACGATTAAGAAAGTGTCTCCAGACTTA 6186  
 RESULT 12  
 ABL05744  
 ID ABL05744 standard; cDNA; 8355 BP.  
 AC  
 AC ABL05744;  
 XX







QY 445 LeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuIleuMetLeu 464  
 Db 5544 -----AACGGCTCTTCATGACGTGGCTGTCGCAATG 5576  
 QY 465 AspProGluValValSerSerCysProAspLeuIleu-----ArgLeuLeuPhe 481  
 Db 5577 GATCCCGAGTGTTCACACT-----CAGCTAATGACGAGCGGCTTCTGTTC 5624  
 QY 482 SerArgAlaGlyGlyLysGlyGlnAlaGlnValProSerPheArgProTyrLeuLeuThr 501  
 Db 5625 TCAAG-----TCTGACAGCGAGTTAGTTGTTCTATCTGTGTTC 5663  
 QY 502 LeuPheThrHisGlnSerSerTrpThrLeuHisGlnCysIleArgValLeuLeuGly 521  
 Db 5664 CTGATCAACCAACCAACGAGGACGACGACGAAAG-----ATGGCGAGTACTCTTC 5720  
 QY 522 LysSerArgGluGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCysIle 541  
 Db 5721 AAGAAATTCATGAGACTACACTACGACCGCTCTCACTGAGAGGACTGACC 5780  
 QY 542 HisValProArgIleTrpGlnGlyArgAspGlnArgThrProGlnLysArgArgGlu 561  
 Db 5781 ACCAATCCGAACTGTGGAAGGACGACCAAGTACATGTCGAAGAAGCTACGCGCGAT 5840  
 QY 562 LeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGlnLeuIle----- 578  
 Db 5841 GCCTTCTTCATGCTGAGGACTTGTGACGAGCGGCTTCCACTTCATCTTCACGA 5900  
 QY 579 ---LeuAlaGlnAlaGlnThrArgSerGlnAspGlyAspTrpAlaAlaCysSerLeuIle 597  
 Db 5901 GGGCTTCCGAGTCCAGTGTGACGACCAAACTATGATTCACCTGCTGCTCG----- 5954  
 QY 598 GlnAlaArgLeuProLeuLeuLeuSerCysCysGlyAspAspGlnSerValArgLys 617  
 Db 5955 -----CGATGACCTTACTGTTTAAGCTCACAGAAAGCGACGACCTCATGTGCAAG 6008  
 QY 618 ValThrGlnHisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArgArg 637  
 Db 6009 GTAATGGAGCAGCTGAG-----AAAGTTCGCTGTCGCACTACTG 6050  
 QY 638 CysAlaGlyAspLeuLeuGlnLeuTyrLeuGlnArgProGlnLeuArgValProValPro 657  
 Db 6051 AAACGTGACGTCCTCCAGAGATGTACATCATGATCCGCGCATCAAGTTCTGAAACCG 6110  
 QY 658 -----GluValLeuLeuHisSerGlnGlyAlaAlaLaserSerSer 670  
 Db 6111 GGCAAAACCGGCGACGAGCGGTACAAATTCGACAACTGAAAGGC----- 6155  
 QY 671 ValCysLysLeuAspGlyLeuIleHisArgPheIleThrLeuLeuAlaAspTrpSerAsp 690  
 Db 6156 ---TCCAGGCGGACAGAGGTGTCCAAACATGTGATCAGCTGTTGGCGAGTGTGTGGCG 6212  
 QY 691 SerArgAlaLeuGlnAsnArgGlyAlaAspAlaSerMetAlaCysArgLysLeuAlaVal 710  
 Db 6213 AAAAAGGACTTGGAAACCTCTCCAGCGACCGACGAGCTCTCTTCGCAAGCTGGCTGCC 6272  
 QY 711 AlaHisProLeuLeuLeuLeuLeuArgHisLeuProMetIleAlaAlaLeuLeuHisGlyArg 730  
 Db 6273 TCCACCGCGCTCTCTCTCTGCGACGCGGCGCTGTCTGTCATTAATGCAAGGTCCG 6332  
 QY 721 ThrHisLeuAsnPheGlnIlePheArgGlnGlnAlaHisLeuSerCysPheLeuHisVal 750  
 Db 6333 GCCCAGCTCAGCATGAAAGCCCTGCGCAGAGCAGACACACTTCCACGATTTTACAGATC 6392  
 QY 751 LeuGlyLeuLeuGlnLeuLeuGlnProHisValPhe-----ArgSerGlnHis 766  
 Db 6393 CTGAGAGAGCTGTGAGCTGTGCAACCACTTTGAGAGAGGCTACAGAAGACGATC 6452  
 QY 767 GlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgGlyLeuLeuAsnTrpArgLys 786  
 Db 6453 CAAAACACGCTG---TGTGCTACTTCACTTCTTCAAG-----CACACACAC 6497

QY 787 SerSerArgHisLeuAlaAlaPheIleAsnLysPheValGlnPheIleHisLysTrpIle 806  
 Db 6498 AACGTAAAGAGCCCTGCCAAATGCTGAACAGTTTGTGACAGATCTCCAAACCTACATC 6557  
 QY 807 ThrTyrAsnAlaProAlaAlaIleSerPheLeuGlnLysHisAlaAspProLeuHisasp 826  
 Db 6558 AACGTCAATCCCTCAAGTCTCTGCTCTCAATCGAAGACGATCGTGGCATCTGTGAAGAA 6617  
 QY 827 ThrPheAsnAspAsnSerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeuSerLeu 846  
 Db 6618 CTTCGACCCAACTACACCTCTACAGCAACTGCAAGGTTTGTGTCCAGGCCCTTCCTCTG 6677  
 QY 847 ProSerArgAspAspArgThrAspArgGlyLeuAspGlnGly----- 861  
 Db 6678 CTGCAGCACAATGCTCCACTCGCGCAGAGAAATTGAGACAGAGAAATTAAAGTACGATAC 6737  
 QY 861 ----- 861  
 Db 6738 GATCTGATGAGCATTTGATGATTAAGCCATCGCCAGCAGACCCGTTGTAACAGAGAT 6797  
 QY 862 -----GlnGlnGlnSerSerAlaGlySerLeuPro 871  
 Db 6798 CCCATCGAAGTAAATCCGAAACACCCATCGATCCAGCAGCAGTAGGGCTCTTATTCG 6857  
 QY 872 LeuValSerValSerLeuPheThrProLeuThrAlaAlaGlnMetAlaProTyr----- 889  
 Db 6858 GTTCAACCTTGCGCTGACAGCGCGGTGAGCTACAGGACATATTCGCGCATCTCTTCG 6917  
 QY 890 -----MetLysArgLeuSerArgGlyGlnThrValGlnAspLeuLeuValLeuSer 907  
 Db 6918 GATCTGCTCAAGATCATTAAGCGATCAACACGAGGAGCTGTGGTCCCATGACG 6977  
 QY 908 AspIleAspGlnMetSerArgArgArgProGlnIleLeuSerPhePheSerThrAsnLeu 927  
 Db 6978 GAGCTGAGAGTGGCTCACTCCAGAGATTTGTTCATCAACGAGAGCTGTCAAGAGCTG 7037  
 QY 928 GlnArgLeuMetSerSerAlaGlnGlnCysArgAsnLeuAlaPheSerLeuAlaLeu 947  
 Db 7038 CTCACCTTAATATCTCGCGAGTGGCCAGATCCGCTCATGCTCATCTGATGATC 7097  
 QY 948 ArgSerMetGlnAsnSerPro-----SerIleAlaAlaPheLeuProThrPhe 964  
 Db 7098 AGGCATCTGAAACCAACATCCCGCAATTGCGACATCAACCTGTGCAACCTTAACCTTAC 7157  
 QY 965 MetTyrCysLeuGlySerGlnAspPheGluValAlaGlnThrAlaLeuArgAsnLeuPro 984  
 Db 7158 ATTCACTGCTCGGGAGACAGAACTCTGCTGTGGCAGCAGCGCATTTACAAATCTGCCG 7217  
 QY 985 GlnTyrAlaLeuLeuCysGlnGlnHisAlaAlaValLeuLeuHisArgAlaPheLeuVal 1004  
 Db 7218 GAGATGTGCGTCTGTCGAGACACAGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 7277  
 QY 1005 GlyMetTyrGlyGlnMetAspProSerAlaGlnIleSerGlnAlaLeuArgIleLeu 1023  
 Db 7278 GCGTTGAATGCTGCTGCTGCAACACTGCGCAGCAAGATTAGAAAGTTCTCCAGACTCTA 7334  
 RESULT 13  
 AAL01467 standard; cDNA; 484 BP.  
 ID AAL01467  
 AC AAL01467;  
 DT 21-NOV-2001 (first entry)  
 XX Human reproductive system related antigen cDNA SEQ ID NO: 1468.  
 DE Human reproductive system related antigen; reproductive system disorder;  
 KW cancer; gene therapy; ss.  
 RN Human sapiens.  
 OS Homo sapiens.  
 XX Homo sapiens.  
 XX WO200155320-A2.

PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01339.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180528.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
DR

DR P-PSDB: AAM95497.  
 XX Isolated nucleic acid molecule encoding a reproductive system antigen  
 PT is used in preventing, treating or ameliorating a medical condition -  
 XX  
 PS Claim 1; SEQ ID NO 1466; 1297pp + Sequence Listing: English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human reproductive system related antigens. These can be used  
 CC in the prevention and treatment of reproductive system disorders,  
 CC including cancer. The present sequence is a coding sequence of the  
 CC invention.  
 XX  
 XX Sequence 484 BP; 90 A; 161 C; 133 G; 84 T; 16 other:  
 S0  
 Alignment Scores:  
 Pred. No.: 4,25e-35 Length: 484  
 Score: 592.00 Matches: 137  
 Percent Similarity: 85.09% Conservative: 0  
 Best Local Similarity: 85.09% Mismatches: 12  
 Query Match: 11.30% Indels: 14  
 DB: 22 Gaps: 2  
 US-09-929-769-7 (1-1029) x ABL01467 (1-484)  
 QY 812 AAlAlAlIeSerPheLeuGlnLysHISAlaAspProLeuHISAspLeuSerPheAspAsn 831  
 DB 4 GGAGCATCTCTCTCTCTGACAGAGACGCGCGCTCCAGACTGCTCTTCCAGAAC 63  
 QY 832 SerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeuSerLeuProSerArAspAsp 851  
 DB 64 AGTGACCTGGTGATCTCTAAATCCCTCTGAGGGCTCAGCTCCGACGAGRACGAC 123  
 QY 852 ArgThrAspArgGlyLeuAspGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 871  
 DB 124 AGGACGACCGAGGCTGACAGAGAGGCGAGGAGAGAGCTCAGCGGCTCTTGCCC 183  
 QY 872 LeuValSerValSerLeuPheThrProLeuThrAlaAlaGluMetAlaProTyrMetLys 891  
 DB 184 CTGGTCAGCGTCTCTCTGTCACCCCTGACCGGCGGAGATGGCCCTACATGAANA 243  
 QY 892 ArgLeuSerArGlyGlnThrValGlu-----Asp 901  
 DB 244 CGGCTTTCGCGGCGCAACGCTGAGGAGTGAATGATCAGCCCTGTTCAMCCAGCCGAT 303  
 QY 902 LeuLeuGlnValLeuSerAspIleAspGlnMetSerArgArgProGlnIleLeuSer 921  
 DB 304 CTGCTGGAGGKTCTGAGTACATAGACGATGTCCCGGAGACCGAGATCCTGAGAC 363  
 QY 922 PhePhe-SerThrAsnLeuGlnArgLeuMetSer---SerAlaGlnGlnCysArgAs 940  
 DB 364 TTCTTCTTCGACCAACGTCAGCGGCTGATGACTTCGCGCCSARNAAGTGTNGCGSAA 423  
 QY 940 nLeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaLaph 960  
 DB 424 CCTTCCTTAAAGCTT-GCCCTGCGSTC-ATGCAAGAACAGCCCGACATTGACGCGCTTT 481  
 QY 960 e 960  
 DB 482 C 482  
 RESULT 14  
 ABL07340  
 ID ABL07340 standard; cDNA; 4351 BP.  
 XX  
 AC ABL07340;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16502.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.

XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO2001/1042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI. 2001-656860/75.  
 XX  
 P-PSDB: ABB63237.  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 16502; 21pp + Sequence Listing: English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 4351 BP; 1235 A; 981 C; 906 G; 1229 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 8.32e-07 Length: 4351  
 Score: 228.00 Matches: 63  
 Percent Similarity: 45.39% Conservative: 70  
 Best Local Similarity: 21.50% Mismatches: 114  
 Query Match: 4.35% Indels: 46  
 DB: 23 Gaps: 4  
 US-09-929-769-7 (1-1029) x ABL07340 (1-4351)  
 QY 773 CysLeuSerPheIleArgLeuLeuAsnTyrArgLysSerSerArgHisLeuAla 792  
 DB 4 TCCTACTTCACACTTCTTCAG-----CACACAGACAGTAAACAGGCTTGC 51  
 QY 793 AlaPheIleAsnLysPheValGlnPheIleHisLysTyrIleThrTyrAsnAlaProAla 812  
 DB 52 CAATATCGTAACAACTTTGTGTCAGATGCTCCAGGCTCATCACTACATCCCTCAAGT 111  
 QY 813 AlaIleSerPheLeuGlnLysHISAlaAspProLeuHISAspLeuSerPheAspAsnSer 832  
 DB 112 GCTTGCTCTTCATCGAACAGTACGACGCTGAGAGAACTTGCACCAAGTACAC 171  
 QY 833 AspLeuValMetLeuLysSerLeuAlaGlyLeuSerLeuProSerArAspAspArg 852  
 DB 172 TCACTAGGAACAACTGAGTTTGTGTCAGCGCGCTTGCCCTCTCCAGACAAAGTCGAC 231  
 QY 853 ThrAspArgGlyLeuAspGlnGly----- 861  
 DB 232 TCGCGAGAGAAATGAGACGAGAGAGATTAAAGTACGAGATGATGATGATTTTC 291  
 QY 861  
 DB 292 GATGTAAAGCCATGCGCCACGACAGCCCTGTGTAACAGAGATCCATCGAAGTGAATCCG 351

```

QY 862 -----Gluc1luc1userSerAlaglySerleuProleuValSerleu 877
DB 352 CAAMACCCATCGATCCAGACAGACAGAGGCGCCCTTATCGGTTTACCTTGGGCTCG 411
QY 878 PheThProleuThralaIalagIumetalaprotyr-----MettysArgleu 893
DB 412 TACAGCGGTCGACACTACACAGGACATATCGCCGCACTTCTCGATCTGGTCAGATCATR 471
QY 894 SerArgGlyGlnThraValGlnAspleuLeuValleuSerAspIleaspgIumetSer 913
DB 472 AAGCACTCCACACACGAGAGCGTGTCTGGGTCATGAGAGAGTGGAGTGCCTCACT 511
QY 914 ArgArgArgProGlnIleuSerPhePheSerThraAsnleuGlnArgleuMetSer 933
DB 532 TCCAAAGATTTGTGTTTCATCAACAGCTGTTGCAACGACGTCTCAACCTTATATTCG 591
QY 934 AlagIgluc1scysArgAsnleuAlaPheSerleuAlaIleuArgserMetGlnAsnSer 953
DB 592 CCGAGTGGCCACAGATCCGGTCCATCGCTTCATCATCTGATCGCATCTGAAGCACAAT 651
QY 954 Pro-----SerIleAlaAlaPheLeuProThraPheMetTyrCysleuGlySer 970
DB 652 CCCGCAATTCGGACATCAACCTGTGCACCTTAACGCTCACTCATGTCAGTGCCTGGGAG 711
QY 971 GlnAsPheGlnValValGlnThraIleuArgAsnleuProGlnIleuAlaIleuLeu 990
DB 712 GAAACATCTCTGGTGGACGAGCGCCATTCGACAAATCTGCGGAGATGTCGGTGTCTG 771
QY 991 GlnGlnIleAlaIleuValleuLeuHISArgAlaPheLeuValGlyMetTyrGlyGlnMet 1010
DB 772 CAGGAACAGCAATTCACATCTCAACGCGTGCCTTCTCGTGGGCTTGAAGTGTGCTG 831
QY 1011 AspProSerAlaGlnIleSerGlnAlaLeuArgIleu 1023
DB 832 AACACTGGCCACGACATAGAAAGTTCTCCAGACTCTA 870

RESULT 15
AAC11139
ID AAC11139 standard; cDNA; 119 BP.
XX
AC AAC11139;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 15214.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX
OS Homo sapiens.
XX
PN EP103401-A2.
XX
PD 06-SEP-2000.
XX
PE 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 990S-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 15214; 71bp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively

```

identified within the present sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 119 BP; 19 A; 44 C; 32 G; 24 T; 0 other;

#### Alignment Scores:

	2.24e-06	Length:	119
Score:	193.00	Matches:	39
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.68%	Indels:	0
DB:	21	Gaps:	0

US-09-929-769-7 (1-1029) x AAC11139 (1-119)

```

QY 941 LeuAlaPheSerleuAlaLeuArgSerMetGlnAsnSerProSerIleAlaIleuPhe 960
DB 2 CTCGCTTCAGCTGGCCCTGCGCTCCATGCAGAACAGCCAGCATTCACCCCTTTC 61
QY 961 LeuProThraPheMetTyrCysleuGlySerGlnAsPheGlnValValGlnThrala 979
DB 62 CTGCCACGTTCTATGTACTGCTGGGACAGCAGACTTTGAGTGTGTCAGACGGCC 118

```

Search completed: May 13, 2003, 23:42:54  
Job time : 615 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2003, 19:39:08 ; Search time 47 Seconds  
(Without alignments)  
4511.117 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 5241  
Sequence: 1 MHILVHAMVILLTGPPRA.....MDESAQISEALRIILHMEAVM 1029

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	5237	99.9	1377	4	Q9P2A8	Q9P2A8 homo sapien
2	4118	78.6	812	4	Q96D36	Q96D36 homo sapien
3	3554	67.8	698	4	Q8WV40	Q8WV40 homo sapien
4	3442.5	65.7	818	11	Q91201	Q91201 mus musculu
5	3244	61.9	640	4	Q9Y3W8	Q9Y3W8 homo sapien
6	1816	34.6	673	4	Q9BT91	Q9BT91 homo sapien
7	1240.5	23.7	310	4	Q9NTD1	Q9NTD1 homo sapien
8	863	16.5	2042	5	Q9W1C5	Q9W1C5 drosophila
9	713	13.6	162	11	Q9CTF7	Q9CTF7 mus musculu
10	160	3.1	1194	4	Q9H7K4	Q9H7K4 homo sapien
11	156	3.0	2472	16	Q91696	Q91696 pseudomonas
12	154	2.9	1638	2	Q87001	Q87001 pseudomonas
13	149	2.8	1217	10	Q64720	Q64720 arabidopsis
14	145	2.8	1310	4	Q9H7L6	Q9H7L6 homo sapien
15	145	2.8	1546	4	Q8TER5	Q8TER5 homo sapien
16	139	2.7	891	12	Q41442	Q41442 infectious

17	139	2.7	1211	11	Q9R0L3	Q9R0L3 ratu
18	139	2.7	1235	11	Q9R0L5	Q9R0L5 ratu
19	139	2.7	1273	11	Q9R0L4	Q9R0L4 ratu
20	138.5	2.6	1007	4	Q8TEB6	Q8TEB6 homo
21	135.5	2.6	840	11	Q9DC40	Q9DC40 mus
22	135.5	2.6	1883	4	Q96SC5	Q96SC5 homo
23	134.5	2.6	2535	10	Q9S240	Q9S240 arabi
24	134	2.6	878	12	Q91SP3	Q91SP3 infect
25	134	2.6	879	12	Q800X3	Q800X3 infect
26	133	2.5	990	4	Q9BXN3	Q9BXN3 homo
27	132.5	2.5	931	4	Q96YV6	Q96YV6 homo
28	132.5	2.5	934	4	Q9P2B0	Q9P2B0 homo
29	132.5	2.5	1542	4	Q15035	Q15035 homo
30	132.5	2.5	1905	4	Q96SC6	Q96SC6 homo
31	132	2.5	879	12	Q9E7P1	Q9E7P1 infect
32	132	2.5	2444	5	Q9VSK6	Q9VSK6 dros
33	131.5	2.5	882	16	Q8ZHC7	Q8ZHC7 yers
34	131.5	2.5	1386	4	Q75064	Q75064 homo
35	131	2.5	1302	4	Q95155	Q95155 homo
36	130	2.5	879	12	Q80H08	Q80H08 infect
37	129.5	2.5	840	11	Q91VQ3	Q91VQ3 mus
38	129.5	2.5	879	12	Q91IX3	Q91IX3 infect
39	129.5	2.5	4835	5	Q8T5T1	Q8T5T1 giard
40	129	2.5	1114	10	Q8S6D2	Q8S6D2 oryza
41	129	2.5	1626	16	Q9RY77	Q9RY77 deino
42	128.5	2.5	1664	4	Q9BZE5	Q9BZE5 homo
43	128	2.4	879	12	Q82630	Q82630 infect
44	128	2.4	1012	11	Q640Z8	Q640Z8 mus
45	128	2.4	1021	16	Q9RV29	Q9RV29 deino

## ALIGNMENTS

## RESULT 1

ID	Q9P2A8	PRELIMINARY:	PRT:	1377 AA.
AC	Q9P2A8:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	KIAA1440 protein (Fragment).			
GN	KIAA1440.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-BRAIN;			
RX	MEDLINE=20181126; PubMed=10718198;			
RA	Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. XVI.			
RT	The complete sequences of 150 new cDNA clones from brain which code			
RL	for large proteins in vitro."			
RL	DNA Res. 7:65-73(2000)."			
DR	EMBL: AB037861; BAA92678.1;			
FT	NON-TER			
SO	SEQUENCE			
Query Match	99.9%; Score 5237; DB 4; Length 1377;			
Best Local Similarity	99.9%; Pred. No. 0;			
Matches 1028; Conservative	0; Mismatches 1; Indels 0; Gaps 0;			
QY	1 MHILVHAMVILLTGPPRADSEFOALIDIFPEPKPLPTAFVLDTSSEALLLPDWKL 60			
DB	349 MHILVHAMVILLTGPPRADSEFOALIDIFPEPKPLPTAFVLDTSSEALLLPDWKL 408			
QY	61 RHIRESEVRLVDAALODLEPOOLLFVOSFGIPVSMKSLLOFLDOAAVAHQDTLEONIM 120			
DB	409 RHIRESEVRLVDAALODLEPOOLLFVOSFGIPVSMKSLLOFLDOAAVAHQDTLEONIM 468			
QY	121 DKNYMAHVEVQHERGASGGQTFHSLTLTASLPPRSDSTEAPRKSSPEQIGGRIRVGT 180			

Dh	469	DKNMNLVEQNHQRAGSGGQTHSLTLTASLPPRRDSTAPRKSSPDP	IQGGIRGT	528
Qy	181	QRLVLPEDDLAGNFLLQIFPLSPDPWQSSSPRPVALA	IQALGGLARVVOGSEVP	240
Dh	529	QLRVLGPEDDLACNFLLQIFPLSPDPWQSSSPRPVALA	IQALGGLARVVOGSEVP	588
Qy	241	TVRVQLQALATLLSSPHGALVMSMHSNHLFACPLRLQ	LCQYQYRCVPDGTGSSFLKYL	300
Dh	589	TVRVQLQALATLLSSPHGALVMSMHSNHLFACPLRLQ	LCQYQYRCVPDGTGSSFLKYL	648
Qy	301	QMLQMLDSPVEGGPPLRAQLRLMASQASAGRRLSDV	RGSLRLABALAFRODLEV	360
Dh	649	QMLQMLDSPVEGGPPLRAQLRLMASQASAGRRLSDV	RGSLRLABALAFRODLEV	708
Qy	361	RAVATATRSRSEQGSVEPDLISKVLQGLIEKRSNHE	LLTAFFSAPADASFPACK	420
Dh	709	RAVATATRSRSEQGSVEPDLISKVLQGLIEKRSNHE	LLTAFFSAPADASFPACK	768
Qy	421	VVSSLILQEEBPLAGCKPGAGDGSLEAVRLSPSGGLV	YMLMLDPEVSSSGCPDL	480
Dh	769	VVSSLILQEEBPLAGCKPGAGDGSLEAVRLSPSGGLV	YMLMLDPEVSSSGCPDL	828
Qy	481	FSRRKKGQQAQVNSFRPLYLTLETTHOSSNPTLHQC	TRVLLGKSREQRDP	540
Dh	829	FSRRKKGQQAQVNSFRPLYLTLETTHOSSNPTLHQC	TRVLLGKSREQRDP	888
Qy	541	IHVRIWQGRQRPQKRREELVLRVQGPFLSYELIAEA	ETRSODGTACSLIQAR	600
Dh	889	IHVRIWQGRQRPQKRREELVLRVQGPFLSYELIAEA	ETRSODGTACSLIQAR	948
Qy	601	LPLLSSCCGDDSVKRYTEHLSGCIQQMGDSVYGR	CRDILLQYLQRP	660
Dh	949	LPLLSSCCGDDSVKRYTEHLSGCIQQMGDSVYGR	CRDILLQYLQRP	1008
Qy	661	LHSEGAASSSYCKLDGLIHFFTLADTSPRALENR	GADAMACRKLAVAHPL	720
Dh	1009	LHSEGAASSSYCKLDGLIHFFTLADTSPRALENR	GADAMACRKLAVAHPL	1068
Qy	721	PMIALLLHGRTLNLFQEFRRQNLSCFLHVLGLLE	LLOPHVRSEHOGALMD	780
Dh	1069	PMIALLLHGRTLNLFQEFRRQNLSCFLHVLGLLE	LLOPHVRSEHOGALMD	1128
Qy	781	LLNRRKSSRHIAAFINKEVQFIHKYITYNAPAI	ISLQKHADPLDLSDNS	840
Dh	1129	LLNRRKSSRHIAAFINKEVQFIHKYITYNAPAI	ISLQKHADPLDLSDNS	1188
Qy	841	LAGLSLPSRDDRTDGLDEEGEESAGSLPLT	SVSLFPLTAAEMAPYMK	900
Dh	1189	LAGLSLPSRDDRTDGLDEEGEESAGSLPLT	SVSLFPLTAAEMAPYMK	1248
Qy	901	DLEVLSDIDEMSRREILISFSTNLQRLMSAECC	CNNLAFSLALSMQNSP	960
Dh	1249	DLEVLSDIDEMSRREILISFSTNLQRLMSAECC	CNNLAFSLALSMQNSP	1308
Qy	961	LPTFMYCIGSODPEVYOTALRNLP	PEVALCOEIAAVLLRAFLVGYGMD	1020
Dh	1309	LPTFMYCIGSODPEVYOTALRNLP	PEVALCOEIAAVLLRAFLVGYGMD	1368
Qy	1021	RIILHMEAVM	1029	
Dh	1369	RIILHMEAVM	1377	
RESULT 2				
Q96D36	ID	PRELIMINARY;	PRT;	812 AA.
Q96D36	AC	Q96D36;		
DT	01-DEC-2001	(Tremblrel. 19, Created)		
DT	01-DEC-2001	(Tremblrel. 19, Last sequence update)		
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)		
DE	Hypothetical 90.0 kDa protein (Fragment).			
DS	Homo sapiens (Human).			

Query Match	Best Local	Similarity	Score	DB	Length	812:			
Matches	809:	Conservative	1:	Mismatches	0:	Indels	0:	Gaps	0:
QY	220	QOALQOELARVQGSSEVPGITVRVLOALATLTLSSPHGALVMSMRHSFLACPLLRQC	279	3	RQALQOELARVQGSSEVPGITVRVLOALATLTLSSPHGALVMSMRHSFLACPLLRQC	62			
QY	280	QYRCVPODGTGSSFLPKYVLQMLQMLDPSVYEGGRLPRLQRLMLASQASAGRLSDVRG	339	63	QYRCVPODGTGSSFLPKYVLQMLQMLDPSVYEGGRLPRLQRLMLASQASAGRLSDVRG	122			
QY	340	ILRLALAFRODLEVSVTVRAVATLRSGGCSVEPRLISKVLQGLLEVRSPHLEEL	399	123	ILRLALAFRODLEVSVTVRAVATLRSGGCSVEPRLISKVLQGLLEVRSPHLEEL	182			
QY	400	TAFFSATDAASPFPACKPVPVVVSSLLQDEEPLAGKPGADGSLAVRLGSSGLVD	459	183	TAFFSATDAASPFPACKPVPVVVSSLLQDEEPLAGKPGADGSLAVRLGSSGLVD	242			
QY	460	WLEMDPEVYSSCPDLQRLFLFSRKGKQAOVPSRPRLTLFTHQSSWPLTHCIRVL	519	243	WLEMDPEVYSSCPDLQRLFLFSRKGKQAOVPSRPRLTLFTHQSSWPLTHCIRVL	302			
QY	520	LGRSREORFDPASIDFLMACIHVPRIMQGRQRTPOKREELVLYRGPELISVELL	579	303	LGRSREORFDPASIDFLMACIHVPRIMQGRQRTPOKREELVLYRGPELISVELL	362			
QY	580	ABAEFRSODGDTAACSILQARLPRLSSCCGDESVRYKTEHLSCGICQMGDSVIGRCR	639	363	ABAEFRSODGDTAACSILQARLPRLSSCCGDESVRYKTEHLSCGICQMGDSVIGRCR	422			
QY	640	DLLDLYLQREPLRVVPEVVLHSSGASSSVCKLDGLHFRITTLADTSDSRALENRA	699	423	DLLDLYLQREPLRVVPEVVLHSSGASSSVCKLDGLHFRITTLADTSDSRALENRA	482			
QY	700	DASMACRIKLAIVNHPDLLLHNLPMIALHGRHLNPOEFRQONHLSCTFHVJGLELQ	759	483	DASMACRIKLAIVNHPDLLLHNLPMIALHGRHLNPOEFRQONHLSCTFHVJGLELQ	542			
QY	760	HVFRSEHOGALWDCILSTFRLLLNFKSSRHIAAFINKEVQETIHKYITYNAPPAISFLQ	819	543	HVFRSEHOGALWDCILSTFRLLLNFKSSRHIAAFINKEVQETIHKYITYNAPPAISFLQ	602			
QY	820	HADPLHDLSDFNQSDLVMLKSLAGLSLPRDRDRTDGLDEEBEESASGLPVYSFLT	879	603	HADPLHDLSDFNQSDLVMLKSLAGLSLPRDRDRTDGLDEEBEESASGLPVYSFLT	662			
QY	880	PLTAAMAPYMKRLSGQVVEEDLEVLSDIDEMSRPRPILSFFSTNLQRLMSSAECCR	939	663	PLTAAMAPYMKRLSGQVVEEDLEVLSDIDEMSRPRPILSFFSTNLQRLMSSAECCR	722			
QY	940	NLAFLSLARSMQNSPISIAAFPLEMYCLGSODFEVYQALATLRLNPREYALLCOEHAVALH	999	723	NLAFLSLARSMQNSPISIAAFPLEMYCLGSODFEVYQALATLRLNPREYALLCOEHAVALH	782			
QY	1000	RAFLVGMVGQMPSAQISALRIHLHEAVM	1029	783	RAFLVGMVGQMPSAQISALRIHLHEAVM	812			

RESULT 3  
ID 08WV40 PRELIMINARY: PRT: 698 AA.  
AC 08WV40:  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SKIN;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: BC018777; AAH18777.1; -  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 698 AA: 77709 MW; C917828091BF4778 CRC64;

Query Match 67.8%; Score 3554; DB 4; Length 698;  
Best Local Similarity 100.0%; Pred. No. 1.7e-249;  
Matches 698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 RLSDVRGGLRLAEALAEAFRODLEVVSSITVRAVATLRSGECSVEEDLSKYLGILEYR 391  
DB 1 RLSDVRGGLRLAEALAEAFRODLEVVSSITVRAVATLRSGECSVEEDLSKYLGILEYR 60  
QY 392 SPHELELTAFFSATADAAPPACKPVVVSLLQEEEPILAGKPGADGSLAEVRLG 451  
DB 61 SPHELELTAFFSATADAAPPACKPVVVSLLQEEEPILAGKPGADGSLAEVRLG 120  
QY 452 PSSGLLVLMLEMDPEVVSQCPDLRLFLSRKKGGAQVDSFRYLLTLTFHOSWPT 511  
DB 121 PSSGLLVLMLEMDPEVVSQCPDLRLFLSRKKGGAQVDSFRYLLTLTFHOSWPT 180  
QY 512 LHQCIRVLLGKSREORFDSASIDFLMACIHVPRIWOGDQRTPOKRREELVLYVGPPEL 571  
DB 181 LHQCIRVLLGKSREORFDSASIDFLMACIHVPRIWOGDQRTPOKRREELVLYVGPPEL 240  
QY 572 ISLVELLIAEAETRSODGDTAACSLIOARLPLLLSCCGDDESVRKYTEHLSGCIQOWGD 631  
DB 241 ISLVELLIAEAETRSODGDTAACSLIOARLPLLLSCCGDDESVRKYTEHLSGCIQOWGD 300  
QY 632 SVLGRRCRDLLOLYLORELPVPEVVLHSEGAASSVCKIDGLIHRFTLLADTSDS 691  
DB 301 SVLGRRCRDLLOLYLORELPVPEVVLHSEGAASSVCKIDGLIHRFTLLADTSDS 360  
QY 692 RALENKGADASMACRKLAAVAHPDLLLRHLPMTAALHGHTHLNFQFROONHLSCLFVYL 751  
DB 361 RALENKGADASMACRKLAAVAHPDLLLRHLPMTAALHGHTHLNFQFROONHLSCLFVYL 420  
QY 752 GLELELQHPVFSEHOGALMDCILSFIRLLNVRKSSRHIAAFINKFVOFIHKYITYNAP 811  
DB 421 GLELELQHPVFSEHOGALMDCILSFIRLLNVRKSSRHIAAFINKFVOFIHKYITYNAP 480  
QY 812 AATISFLQKHADPLHDLSPNSDLVMLKSLAGLSLPSRDDRTDGLDEGESESSAGSLP 871  
DB 481 AATISFLQKHADPLHDLSPNSDLVMLKSLAGLSLPSRDDRTDGLDEGESESSAGSLP 540  
QY 872 LVSYSLEFPLTAEMAPYKKRLSRGQTVEDLLEVLSDIDEMSRRRPEILSFSTNLQRLM 931  
DB 541 LVSYSLEFPLTAEMAPYKKRLSRGQTVEDLLEVLSDIDEMSRRRPEILSFSTNLQRLM 600  
QY 932 SSAECCRNLAESLALRSQNSPSIAAFLPTFMYCLGSDQFEVYVOTALRNLPETVALLQ 991  
DB 601 SSAECCRNLAESLALRSQNSPSIAAFLPTFMYCLGSDQFEVYVOTALRNLPETVALLQ 660  
QY 992 EHAAYLHRAFLVGMYGMDPSAQISEALRIITLMEAYM 1029  
|||||

DB 661 EHAAYLHRAFLVGMYGMDPSAQISEALRIITLMEAYM 698

RESULT 4  
ID 091Z01 PRELIMINARY: PRT: 818 AA.  
AC 091Z01:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Strausberg R.;  
RA Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: BC010333; AAH10333.1; -  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 818 AA: 90949 MW; A5128C8692BB7663 CRC64;

Query Match 65.7%; Score 3442.5; DB 11; Length 818;  
Best Local Similarity 82.3%; Pred. No. 2.7e-241;  
Matches 673; Conservative 70; Mismatches 72; Indels 3; Gaps 2;

QY 215 VALALQALQOELARVYGQSPVEVGTIVRYLQALATLLSSPHGALVMSHRSFLACPL 274  
DB 1 VALALQALQOELARVYGQSPVEVGTIVRYLQALATLLSSPHGALVMSHRSFLACPL 60  
QY 275 IROLCOYORCPDPTGESSFLKYLLOMLQMDSPVEGGPIPAOLMILASQASARRRS 334  
DB 61 MROLYOYRANVPDPTGESSFLKYLLOMLQMDSPVEGGPIPAOLMILASQASARRRS 120  
QY 335 DVNKGGLRLAEALAEAFRODLEVVSSITVRAVATLRSGECSVEEDLSKYLGILEYRSPH 394  
DB 121 DVNKGGLRLAEALAEAFRODLEVVSSITVRAVATLRSGECSVEEDLSKYLGILEYRSPH 180  
QY 395 LEEELTAFFSATADAAPPACKPVVVSLLQEEEPILAG-KPGADGSLAEVRLGSPS 453  
DB 181 LEEELTAFFSATADAAPPACKPVVVSLLQEEEPILAG-KPGADGSLAEVRLGSPS 240  
QY 454 SGLLVLMLEMDPEVVSQCPDLRLFLSRKKG--GAQVDSFRYLLTLTFHOSWPT 511  
DB 241 SGLLVLMLEMDPEVVSQCPDLRLFLSRKKG--GAQVDSFRYLLTLTFHOSWPT 300  
QY 512 LHQCIRVLLGKSREORFDSASIDFLMACIHVPRIWOGDQRTPOKRREELVLYVGPPEL 571  
DB 301 LHQCIRVLLGKSREORFDSASIDFLMACIHVPRIWOGDQRTPOKRREELVLYVGPPEL 360  
QY 572 ISLVELLIAEAETRSODGDTAACSLIOARLPLLLSCCGDDESVRKYTEHLSGCIQOWGD 631  
DB 361 ISLVELLIAEAETRSODGDTAACSLIOARLPLLLSCCGDDESVRKYTEHLSGCIQOWGD 420  
QY 632 SVLGRRCRDLLOLYLORELPVPEVVLHSEGAASSVCKIDGLIHRFTLLADTSDS 691  
DB 421 SVLGRRCRDLLOLYLORELPVPEVVLHSEGAASSVCKIDGLIHRFTLLADTSDS 480  
QY 692 RALENKGADASMACRKLAAVAHPDLLLRHLPMTAALHGHTHLNFQFROONHLSCLFVYL 751  
DB 481 RSESESVADANAMACRKLAAVAHPDLLLRHLPMTAALHGHTHLNFQFROONHLSCLFVYL 540  
QY 752 GLELELQHPVFSEHOGALMDCILSFIRLLNVRKSSRHIAAFINKFVOFIHKYITYNAP 811  
DB 541 GLELELQHPVFSEHOGALMDCILSFIRLLNVRKSSRHIAAFINKFVOFIHKYITYNAP 600  
QY 812 AATISFLQKHADPLHDLSPNSDLVMLKSLAGLSLPSRDDRTDGLDEGESESSAGSLP 871  
DB 601 AATISFLQKHADPLHDLSPNSDLVMLKSLAGLSLPSRDDRTDGLDEGESESSAGSLP 660  
QY 872 LVSYSLEFPLTAEMAPYKKRLSRGQTVEDLLEVLSDIDEMSRRRPEILSFSTNLQRLM 931  
|||||

```

Db 661 LVASVLSITPLVADVAPHPMKRLSGRAVEDLETLSIDDEMSRRRPEVLGFSTNLQRLM 720
QY 932 SSAECCRNIAFLSLALRSMQNSPSIAAFLPTFMVCLGSDQDEEVYOTALRNLPEVALLCQ 991
Db 721 SSAECCRNIAFLSLALRSMQNSPSIAAFLPTFMVCLGSDQDEEVYOTALRNLPEVALLCQ 780
QY 992 EHAVALLHRAFLVGMVGMQDPSAQISEALRIILMEAVM 1029
Db 781 EHAVALLHRAFLVGMVGMQDPSAQISEALRIILMEAVM 818

```

## RESULT 5

```

Q9Y3W8 09Y3W8 PRELIMINARY; PRT; 640 AA.
AC Q9Y3W8;
DT 01-NOV-1999 (TReMBrel. 12, Created)
DT 01-NOV-1999 (TReMBrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBrel. 19, Last annotation update)
DE Hypothetical 71.4 kDa protein (fragment).
GN DAFZP586J0619.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Wambolt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL050110; CAB43278.1; -
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 640 AA: 71375 MW: 1AC887FC2075B21A CRC64;

```

```

Query Match 61.9%; Score 3244; DB 4; Length 640;
Best Local Similarity 100.0%; Pred. No. 5e-227;
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 397 ELTAFSSATADAAAPPACKPVVVVSSLLIQEEPLAGKPGADGSLAVRLGSSGL 456
Db 8 ELTAFSSATADAAAPPACKPVVVVSSLLIQEEPLAGKPGADGSLAVRLGSSGL 67
QY 457 LVDMLEMLDEPVVSSCDLQRLIFSRKKGGOVPSFRYLLTLTHOSSWPTLQCI 516
Db 68 LVDMLEMLDEPVVSSCDLQRLIFSRKKGGOVPSFRYLLTLTHOSSWPTLQCI 127
QY 517 RVLLGKREORFDPASLIDFLMACIHVPRIWGRDQRTPKRREELVLRVQGPPELLISLVE 576
Db 128 RVLLGKREORFDPASLIDFLMACIHVPRIWGRDQRTPKRREELVLRVQGPPELLISLVE 187
QY 577 LIIAEATRSQDDGTAAACSLIOARLPLLLSCCGDDSVKRVTEHLSGCIQOWGDSVYGR 636
Db 188 LIIAEATRSQDDGTAAACSLIOARLPLLLSCCGDDSVKRVTEHLSGCIQOWGDSVYGR 247
QY 637 RCRDLILQLLOPELRLVPEVLLHSEGAASSVCKLDGLIHFTLLADTSDSRALEN 696
Db 248 RCRDLILQLLOPELRLVPEVLLHSEGAASSVCKLDGLIHFTLLADTSDSRALEN 307
QY 697 RGADASMACRKLAVAHPLLLRHLPMTAALLHGRTHLNFOEFROONHLSGFLVHGLLEL 756
Db 308 RGADASMACRKLAVAHPLLLRHLPMTAALLHGRTHLNFOEFROONHLSGFLVHGLLEL 367
QY 757 LQPHVFRSEHOGALMDCLSFIRLLLNYSRRHLAAFINKEVOFHRYTYNAPAAISF 816
Db 368 LQPHVFRSEHOGALMDCLSFIRLLLNYSRRHLAAFINKEVOFHRYTYNAPAAISF 427
QY 817 LQKHADPLHDSFDNSDLVWLKSLAGLSLPSRDDRTDRDLDEGEESAGSLPLVSVS 876
Db 428 LQKHADPLHDSFDNSDLVWLKSLAGLSLPSRDDRTDRDLDEGEESAGSLPLVSVS 487
QY 877 LFTPLTAENAPYMKRLSRGQTVEDLEVLSDIDEMSRRPPELLISFSTNLQRLMSAAE 936

```

```

Db 488 LFTPLTAENAPYMKRLSRGQTVEDLEVLSDIDEMSRRPPELLISFSTNLQRLMSAAE 547
QY 937 CCRNLAFSLALRSMQNSPSIAAFLPTFMVCLGSDQDEEVYOTALRNLPEVALLCQ 996
Db 548 CCRNLAFSLALRSMQNSPSIAAFLPTFMVCLGSDQDEEVYOTALRNLPEVALLCQ 607
QY 997 LLHRAFLVGMVGMQDPSAQISEALRIILMEAVM 1029
Db 608 LLHRAFLVGMVGMQDPSAQISEALRIILMEAVM 640

```

## RESULT 6

```

Q9BT91 09BT91 PRELIMINARY; PRT; 673 AA.
AC Q9BT91;
DT 01-JUN-2001 (TReMBrel. 17, Created)
DT 01-JUN-2001 (TReMBrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBrel. 19, Last annotation update)
DE Hypothetical 75.1 kDa protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC004286; AAH04286.1; -
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 673 AA: 75055 MW: 6CE17654F5B3AB3DE CRC64;

```

```

Query Match 34.6%; Score 1816; DB 4; Length 673;
Best Local Similarity 99.7%; Pred. No. 2.6e-123;
Matches 360; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 MHILVHAMVILTLTGPRADDSFQALDIWPEEKPLPTAFIVTSEALLIPDWLKL 60
Db 304 MHILVHAMVILTLTGPRADDSFQALDIWPEEKPLPTAFIVTSEALLIPDWLKL 363
QY 61 RMIRSEVRLVDAADOLEFOOILLFVOSFGIPVSSSKLLOFDAVAHDPOTEGNIM 120
Db 364 RMIRSEVRLVDAADOLEFOOILLFVOSFGIPVSSSKLLOFDAVAHDPOTEGNIM 423
QY 121 DKNYMAHVEVQHERGASGQTFHSLLTASLPPRRDSTEAPKPKSSPEQPIGGRIRVGT 180
Db 424 DKNYMAHVEVQHERGASGQTFHSLLTASLPPRRDSTEAPKPKSSPEQPIGGRIRVGT 483
QY 181 QLRVLGPEDDLACMFLOIFPLSDPPRQSSSPRVALALQOALGOELARVQGSPEVPGI 240
Db 484 QLRVLGPEDDLACMFLOIFPLSDPPRQSSSPRVALALQOALGOELARVQGSPEVPGI 543
QY 241 TVRYLQALATLLSPHGGALVMSWHRSHFLACPLRLQCYORCVPODTGFSSLFKVLV 300
Db 544 TVRYLQALATLLSPHGGALVMSWHRSHFLACPLRLQCYORCVPODTGFSSLFKVLV 603
QY 301 QMLQWLDSPEVEGPIRAQRLMLASQASAGRRLSDVNGGLRLAELARQODEVVSSTV 360
Db 604 QMLQWLDSPEVEGPIRAQRLMLASQASAGRRLSDVNGGLRLAELARQODEVVSSTV 663
QY 361 R 361
Db 664 R 664

```

## RESULT 7

```

Q9NTD1 09NTD1 PRELIMINARY; PRT; 310 AA.
AC Q9NTD1;
DT 01-OCT-2000 (TReMBrel. 15, Created)
DT 01-OCT-2000 (TReMBrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBrel. 15, Last annotation update)

```



DE Hypothetical 34.7 kDa protein (Fragment).  
 GN DKFZP34C0126.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Ansoer W., Wirtner U., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AL3137358; CAB/0710.1;  
 KW Hypothetical protein.  
 FT NON-TER 1  
 SO SEQUENCE 310 AA; 34728 MW; EDBA23F1D0C08521 CRC64;  
 Query Match 23.7%; Score 1240.5; DB 4; Length 310;  
 Best Local Similarity 92.6%; Pred. No. 5.2e-82;  
 Matches 249; Conservative 2; Mismatches 7; Indels 11; Gaps 1;  
 QY 756 LLOPHVSEHOGALMDLSPFIRLLNVRKSSRLAFLINKFQFIHXYTNAPATLS 815  
 Db 1 LLOPHVSEHOGALMDLSPFIRLLNVRKSSRLAFLINKFQFIHXYTNAPATLS 60  
 QY 816 FLOKHADPLHDLSPDNDLVMIKSLAGISLPSRDDRTDGLDEGEESAGSLPLSV 875  
 Db 61 FLOKHADPLHDLSPDNDLVMIKSLAVLSLPSRDDRTDGLDEGEESAGSLPLSV 120  
 QY 876 SLFPPPLTAEMAPYKRLSRGQTVEDLLEVLSDIDEMSRRPETLSFTNQLRMSSAE 935  
 Db 121 SLFPPPLTAEMAPYKRLSRGQTVEDLLEVLSDIDEMSRRPETLSFTNQLRMSSAE 180  
 QY 936 ECCRRLAASLRSNONSPTIAAFLPTMVCIGSGDEEVQATLRNLPVALCOENAA 995  
 Db 181 ECCRRLAASLRSNONSPTIAAFLPTMVCIGSGDEEVQATLRNLPVALCOENAA 240  
 QY 996 VLLHRAFLVG-----MYGMDPS 1013  
 Db 241 VLLHRAFLVG-----MYGMDPS 269  
 RESULT 8  
 Q9W1C5 PRELIMINARY; PRT; 2042 AA.  
 ID Q9W1C5  
 AC Q9W1C5  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
 DE CG3173 protein.  
 GN CG3173.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Bokorova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferltera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulyov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.D., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Ventle E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of Drosophila melanogaster."  
 RT Science 287:2185-2195(2000).  
 DR EMBL: AE003462; AAF47145.1;  
 RL Flybase: FBgn0034964; CG3173.  
 SO SEQUENCE 2042 AA; 233746 MW; 9144DB7B51F52E5 CRC64;  
 Query Match 16.5%; Score 863; DB 5; Length 2042;  
 Best Local Similarity 25.4%; Pred. No. 2e-53;  
 Matches 280; Conservative 191; Mismatches 402; Indels 230; Gaps 32;  
 QY 1 MHILVHVMVILTTGPRPADSEFOALDIDFPEKPLPTFPAIVDTSEAL-LIPDMK 59  
 Db 1084 LHINIITHAFILTTLS--NSNPESIPILDYFPFGRPPAVFLPSMPOEQYQLPDMK 1141  
 QY 60 LMRSEVLRVDAALQDLEPOQLLFVQSGFIPVSSMSKLLQFLDQAAVHDPQTLQONI 119  
 Db 1142 LKMIRSSVDRLEALINDLTPQIVLFVQNFQVPSMSKLLAMDITAVLEQFDLVKNKI 1201  
 QY 120 MDKNTMAHLYEVOHERGASGQTFHSLTASLPPRRDSTLEAKPKSSPPQPGGRIKRG 179  
 Db 1202 LMKAVLAQLEITQARGAKNG---HYVQALDLHSHQVTPDLKITS---VVIQEAVER- 1254  
 QY 180 TQLRVLGPEPDLAGMFLQIFPLSPDRMOSSSPRPALALQALGQELARVVGSPPEVG 239  
 Db 1255 -----DDYDS-----SDSDRPNPLFATEVQAQITITQDQLTESHS 1291  
 QY 240 ITVRVLAQATLTLSSPH-----GGALVMSMRSHFLACPLLRQLQCYQR 283  
 Db 1292 DCRSLIQKLDMLASPNRSRAVDVNAITEVLAVGCVTMSRH-----ACTFLR----- 1339  
 QY 284 CVPQDTGFSFLKYLVLQMLQMLDSGVGCGPLRAQRLYLAQASAGRLSDVRGGLRL 343  
 Db 1340 -----TFSCMLHSDKYHLE-----NALQKNLSMF-----KHTPADSSLQK 1377  
 QY 344 AEALAFRODLEVVSTVRVAVIATLRSGECCSVEPDLIS--KVILGLEIERSPHLEELTA 401  
 Db 1378 SE-LYHESVLVMLNRSRIYA-----QQFRKNTALVAKKRIVRAVQ----- 1418  
 QY 402 FFSATADAAPPACKPVVAVSSLLQEEEPAGCGPGADGSGLEAVRLGPSSGLLVML 461  
 Db 1419 SFDQTKDS-----KTVAKSKSDQLFH-----NGLEFDLWL 1447  
 QY 462 EMLDEVVSSCPDQL--RLFSRKKGQAQVPSFRYLLTLTTHOSSWPTLHOCITAV 518  
 Db 1448 SEMDEIVST--QLMKRFLESK-----SCSEFRYLLSLNHQTNWPTIER-TAE 1495  
 QY 519 LIGKSEGEQFDPASDIFLMACIHPRIVWGQDQRTPPKREELVRYVGPPLISLVELI 578  
 Db 1496 YLFKNFHEVDYAVYLVNTEALTNPKLMKGKDKYMSKNVRDAPFMILMTSELEPSHPT 1555  
 QY 579 ---LAETRSQDGDGTACSLIQARLPLLLSCCGGDESVKRYTEHLSGCIQMGDSVL 634

```

Db 1556 LHEGLSEVKDKSKNDYDKLCS-----RMNLLFKLEKRRDLVKVKEHVE-----KSSVS 1605
QY 635 GRRCRDLLDLYLQRPRLRVVPEVLLHSEGAASSV-----CKDGLIHRTLLA 686
Db 1606 DYKLQYVLOQMYNYPRIKFLP-----CKTGEQAYKIQNLKGCQADKYSNNLITCLG 1658
QY 687 DTSDSRALENGDASNAACKRLAVAPHLRLRLPMIALHGHRTLNPFQFRQONHLS 746
Db 1659 SLVKKPDEFELSTDETLRLKRLASHPLFLRQLGVLSIMOGRAQLSMKRLREHNHR 1718
QY 747 FLVHLGLLELLQPHVF-----RSEHGALMDCLLSFIRLLNRYKSSRHIAFIKQFOFI 802
Db 1719 FVQILRLLELQPTIFEBAYKNETQNTL-SCYENFER-----HSSNVEACOMLKKFVOML 1773
QY 803 HKYITVAPAPAISEFLQKHAFLDHLSPDNSDLYMLKSLAGLSLPSRDDFTDRGDEGC- 861
Db 1774 QAYINYPSSALLFEQYVGLKELAKYISLGLQYLVQAVALLQKHSATLEDDDEV 1833
QY 862 -----EESSAGSLPLVSVSLFTPLTAEMA 887
Db 1834 KYEYDLDEHDPKPSAKRPVTEDEPIEVNPTQPIDSSSRGSLSVLFLGYSRSNYTDIS 1893
QY 888 PY-----MKRLSRGQTVEDLLEVLSDIDMSRRPEILSFSTNLQRLMSSAECCRNIAF 943
Db 1894 PHFLDLVKIIOSTEDVYLGPMOELCTSKRFVFTNELFERLLNLFSPSAQIRSTAF 1953
QY 944 SLALRSQNSP---SIAAFLPTMYCIGSODFEVOTALRNLEPEYALCOEHAAYLHR 1000
Db 1954 IILIRHLKHNQNSDINCLTNINAVIQCLRDENSSVAATAIDNLEMSVLLQEHMIDITV 2013
QY 1001 AFLVGMGQMDPSAQISEALRIL 1023
Db 2014 AFSGLKSCINTGHQIRKVLQTL 2036

RESULT 9
Q9CTF7 AC Q9CTF7 PRELIMINARY; PRT; 162 AA.
ID Q9CTF7 AC Q9CTF7
AC Q9CTF7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 1110015K06Rik protein (Fragment).
GN 1110015K06Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Kono H., Baldarelli R., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).

```

```

DR EMBL: AK003728; BAB22963.1; -
DR MGD; MGI:1915760; 1110015K06Rik.
FT NON TER
SO SEQUENCE 162 AA; 17939 MW; E343A7CF040CF7A CRC64;
Query Match 13.6%; Score 713; DB 11; Length 162;
Best Local Similarity 87.0%; Pred. No. 4.1e-44;
Matches 141; Conservative 12; Mismatches 9; Indels 0; Gaps 0;
QY 868 GSLPLVSVSLFTPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDMSRRPEILSFSTNL 927
Db 1 GSLPLVSVSLFTPLTAEMAPYMKRLSRGQTVEDLLEVLSDIDMSRRPEILSFSTNL 60
QY 928 QRLMSAECCRNIAFLALRSQNSPISIAAFLPTMYCIGSODFEVOTALRNLEPEYA 987
Db 61 QRLMSAECCRNIAFLALRSQNSPISIAAFLPTMYCIGSODFEVOTALRNLEPEYT 120
QY 988 LLCOEHAAYLHRAFLVGMGQMDPSAQISEALRILMEAVM 1029
Db 121 LLCOEHAAYLHRAFLVGMGQMDPSAQISEALRILMEAVM 162

RESULT 10
Q9H7K4 AC Q9H7K4 PRELIMINARY; PRT; 1194 AA.
ID Q9H7K4 AC Q9H7K4
AC Q9H7K4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-JUN-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FLJ00068 protein (Fragment).
GN FLJ00068.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLN;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK024475; BAB15765.1; -
DR InterPro: IPR001251; GRAL-TRIO.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhGEF.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhGEF; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhGEF; 1.
DR SMART: SM00516; SEC14; 1.
DR PROSITE: PSS0003; PH_DOMAIN; 1.
FT NON TER
SO SEQUENCE 1194 AA; 131087 MW; 86Z01A3A0FEACB67 CRC64;
Query Match 3.1%; Score 160; DB 4; Length 1194;
Best Local Similarity 20.9%; Pred. No. 0.01;
Matches 229; Conservative 104; Mismatches 372; Indels 390; Gaps 47;
QY 74 ALDLEPQQLLFVQSGFIPVSSMSKLLQFLDAVADHPQTLF-QNIMDKN----- 123
Db 58 AYDDEELQ-----GSLPSKRFOLPRADEGDAGRGVESSVLSLEGPGSSVEYL 108
QY 124 ---YMAHLVEYHGERGASGGQTFHSLTLASLPPRRDSTEAPKPSSEOPRIGGRIRVGT 180
Db 109 LCPMSHLSLAQGSDFPGV-----LVGDPGPERAMPSPGSLPGLSDPVLG----- 157
QY 181 QLRVLGPEDLACFLQIFLPSPPKQSSPPRYVALALQALQGLARVYQSPPEYPT 240
Db 158 ---DPLSESKL-LEAASG-----SGLPKPADCLADLQWEL--LASGATLPG-- 202
QY 241 TVRVQLATLTLSSPHGALVMSHRSHPACLP-----LRLQLOVQRCV 285

```

```

Db 203 -TRVOGRAVILLCAHSPAMLOSCSOELLRLILRLSIPREYVALGLTVLPARICA 261
QY 286 PDGTFSSSL-----FLKVLQMLQWLDSPGVESGLPAOLRMLASOA-----SA 329
Db 262 PSSSLFSGLSQLEAPGAVYVL-----VSTLKEVPSGLQLEQLPSLSLHIPIPA 316
QY 330 GRLSLDVROGLRLAEA-LAFRODELVSSVRAVATLRSGEQCSVEPDLISKYLOGLI 388
Db 317 GLPFS-LGGGLPYCHQAMLDFFRRRLLEALLQNCQAMCA-----LLQGA1 358
QY 369 EYRSHLEELLAFSADADAAPPACKPVVYVSSLLQEEPLAGCPGADGSGLEAV 448
Db 359 E-----SVKAVPQM-----EPG----- 371
QY 449 RLGPSSGLVDMLEMDPEVVSQCDLQRLFLFSRKKKGQAQVPSFRYLLTLFTHOSS 508
Db 372 -EVGQLLOQTEVLMQOVDS-----PWLML----- 396
QY 509 WPTLHOCIRVLLGKSREORFDPASLDFLMACIHVPRIWGHODORTPOKRREELVRYOG 568
Db 397 -QC-----QGGRELTWLRKOEVPVETLSPDYRTAMOKADELYDRVDS 436
QY 569 PELISVVELLAEATRSODGDTACSLQALPLLLSCCDDSDSVKRYTHLSGCTIO 628
Db 437 -LHQLTLQSNQRIQ-----ALELVOTL-----EARESGLHQIEWMLQ 474
QY 629 WGDVYL-----GRRCRDLLLOLYLQREPLRVPEVLLHSE-----GAASSVCKLDGLIHRF 681
Db 475 VGMPLERAGEPSLDMLOAGSFOELYOVAGROGKFLQPLTGWAELIDPPGANF 534
QY 662 ITLLAD-TSDSRALENRGADASMACRIKAVANPRLRLRLPMAIALHGRTHLNQEFRO 740
Db 535 LALRAQLTEFSALQOR-----CQRLADARLFLQ-----FREALTVNHEGO 576
QY 741 QNHLSCLFVHVLGILLDQPHVRSE-----HOGALM-----DC 773
Db 577 R-----VLALEQERGGVVLQDLQHLWTRHPLRPAHFRKMAALATGSEAIROEC 628
QY 774 LLSFRLLLNYKSSRHAAFINKEVFOITNYNAPAIIFLOKHADPLHD-----LS 828
Db 629 RWAMARCODTWALDQKLEASL-----KLPRVGSSTSLCVSYPARAPRPLKAYS 680
QY 829 FDNSSLVWKSLAGLSLPSRDRTDGLDEGEBSASGLPLVSVSLFTP----- 880
Db 681 FD-----RNIGOSLSEFACHCHAAITAACRPEAGGALPQASPTVPPGSSDPSRL 733
QY 881 -LTAEMA-----PYMKRLSRGQTVEDLEVLSDIDMSRRRPEILSFEFTNIQRL- 930
Db 734 NRIQVLAEMWATEREYVRALE--YTMENYFPELLRPDPVQIGORAHLPFC-NLEKRD 790
QY 931 -MSSAECCRNLAFLSLARSMONSPSTAAFL-----PTFMCLGSSQD----- 972
Db 791 FHCHEFLMELEACTR-----HPRVAVAFLRRHVQFGMYALXSKKPPRSALM 838
QY 973 -FEVOTAL-----RNLPEYALLQEHAAVILLHRAFLVGMYGOMD 1011
Db 839 SSYGHTEFRKDKQALDHLDLASYLKPIQRMGKVALLOE-----LAAA-----CGGPTQ 889
QY 1012 PSAOISEALRLTME 1026
Db 880 ELSALREKOSLVHQ 904

```

```

RESULT 11
Q91696 PRELIMINARY: PRT: 2472 AA.
AC Q91696
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Still frameshift probable component of chemotactic signal transduction
system.
GN PA0413.

```

```

OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxId=287;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.J., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RT Nature 406:959-964(2000).
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
CC -1- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
CC EMBL: AE004478; AAG03802.1; -.
DR HSSP: O56310; 1B30.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR004358; Bact_sens_pr_C.
DR InterPro: IPR002545; Chem.
DR InterPro: IPR004359; HIS_KIN_sig.
DR InterPro: IPR001789; Response_reg.
DR Pfam: PF01584; Chem; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR Pfam: PF01627; Hpt; 6.
DR Pfam: PF00072; response_reg; 1.
DR PRINTS: PR00344; BCTRLSENSOR.
DR ProDom: PD000039; Response_reg; 1.
DR ProDom: PD003142; Hpt; 1.
DR SMART: SM00260; Chem; 1.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00073; HPT; 5.
DR SMART: SM00448; REC; 1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase;
KW complete proteome.
SQ
SEQUENCE 2472 AA; 268573 MW; 9886D8A88417A5E CRC64;
Query Match 3.0%; Score 156; DB 16; Length 2472;
Best Local Similarity 19.1%; Pred. No. 0.056;
Matches 229; Conservative 142; Mismatches 386; Indels 440; Gaps 53;

```

Db 1150 PGSDALLERIEALHGERIASLEAERYSDAGERRPDLLEAFVBCMDILLDAEDLLERWH 1209  
 QY 345 EALAFRODLEVSSTVAATATLRSQ-----ECSVEPDLISKVLOGLI----- 388  
 Db 1210 EHPEROEL-----SALREELSTLDGRARHAEIPQVEELCAALLALAYAVEEGRLAVSPAF 1265  
 QY 389 --EVRSPH-----LEELITA-----FFSATDAASAPP 414  
 Db 1266 FEEARQAHREALIGMDVAAGLOVTPPERVAALQELLLEAPAAEAVFIPDESIGADDFP 1325  
 QY 415 -----ACKPVVVVSSLLQEEEPPLAGKPGADGSL-----BAVRLGPSG-LLV 458  
 Db 1326 PEDDEPALPEAVEEAGAPAEETVPAPAPAPGHEDEEMVSTFLEAVDILSAGALA 1385  
 QY 459 DWLEMDPEVYSSCPDQLRLFLSRRKKGQAOVPSFRPYLLTLFHQSSMPTLHOCIRV 518  
 Db 1386 QW-QAEPGALSSLSALQ-RDLHTLKGARMAEIAEIGD-----LAHE-----LEA 1428  
 QY 519 LIGKSRQRPDPSASLDFLWACHIVPRIMOGROQT---PQKRREELVLRV---QGP-- 569  
 Db 1429 LYEGVDRRYQHSPQLAGLLOACH-DRLAQDLQDSAGQPLADPHDLQSIIRFRQGPVA 1487  
 QY 570 -----ELISVELLIEAETRSQDGTAAQSLQARPLPILSCCGDDSVKRVTEHLS 623  
 Db 1488 EAATPGAESPEVELVAPA---VEEPAAPAEAEFEEDPELVEITL---EEGFDILDSAA 1541  
 QY 624 GCIOQWGDVY----- 633  
 Db 1542 AALQRMWDVDNTELEALQDRLTLKGARMAEIGEIGDLAHEFLYEGLCGRILAS 1601  
 QY 634 -----LGRRCRDLLQY-----LQR---PELAVPYPEVL-----LH 662  
 Db 1602 PALEGLRQCHDELAEMLEAVRGHRTLPGQALIAETIRLRSDDEQLSVPTSLKPLA 1661  
 QY 663 SEGASSSVYCKLDGLIRFTILLAD-----TSDRALENRQADSMACRLAVA 711  
 Db 1662 AKGAADSESLDIFLEADADDLELELALGRMDGNGDQPLDD----- 1706  
 QY 712 HPULLHRLPMIALHGRHNLNFOEFRQONHLSCLHVLGLLELLOPHVFRSEHGALW 771  
 Db 1707 --LIRLHTLKGARLAGQTELG-----NLADHLEOHHTLDAQOQCAPV 1747  
 QY 772 -DCLLSFIRLLNVRKSSRHIAFINKEVQFIHKYITYNAPPAISFLOKHADPLHDSFD 830  
 Db 1748 PDSL-----LDQSGLEGLOVOYDLRELRLE 1775  
 QY 831 NSDLVMLKSLAG-----LSLPSRDDTRD---GLDEGEESSSAGSLPLVSVSLFTP 880  
 Db 1776 DDE-----AGRPPEPAQALVQADD-TDRAVASALAEELRLAPAGA--IMAAEAPR 1824  
 QY 881 LTAAEAPYMKRSLRGQTEVDELLVSLDIDMSRRP-ELISFSTVLQRLMSAE 936  
 Db 1825 AABATTLPEVRKAQEAQ-----EASRRAPQELVKPAELLENLVNLAGE 1870

RESULT 12  
 087001 PRELIMINARY; PRT; 1638 AA.  
 AC 087001: 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE ChpA.  
 GN ChpA.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC WHITCHURCH C.B., Young M.D., Hobbs M., Mattick J.S.,  
 RA "Pseudomonas aeruginosa chemotactic transduction genes pilL, chpA chpB

RT and downstream genes chpC, chpD and chpE.";  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 CC - SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE  
 CC KINASES.  
 DR EMBL: U79580; AAC23931.1; -  
 DR HSSP: Q56310; 1B30.  
 DR InterPro: IPR003594; ATPbind\_ATPase.  
 DR InterPro: IPR002545; Chew.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR002570; Hpt.  
 DR InterPro: IPR001789; Response\_reg.  
 DR Pfam: PF01584; Chew. 1.  
 DR Pfam: PF02518; HATPase\_c. 1.  
 DR Pfam: PF01627; Hpt. 6.  
 DR Pfam: PF00072; response\_reg. 1.  
 DR Prodom: PD000039; Response\_reg. 1.  
 DR Prodom: PD003142; Hpt. 1.  
 DR SMART: SM00260; Chew. 1.  
 DR SMART: SM00387; HATPase\_c. 1.  
 DR SMART: SM00073; Hpt. 5.  
 DR SMART: SM00448; REC. 1.  
 KW Kinase; Phosphorylation; Sensory transduction; Transferase.  
 SQ SEQUENCE 1638 AA; 178594 MW; 1C760B1869E2381E CRC64;

Query Match 2.98; Score 154; DB 2; Length 1638;  
 Best Local Similarity 19.18; Pred. No. 0.044;  
 Matches 229; Conservative 141; Mismatches 387; Indels 440; Gaps 53;

QY 8 AMVLTLPGRPADDSFQALDIPPEKPLPAFLVDNSE--EAL--LLPDV----- 57  
 Db 12 AONVPVSLPPPADE-----PYDELRVFTIEEGEVLTEGRIPLPAKADHD 61  
 QY 58 -----LK-LRMIRSEVL-----RLVDAALQDLPEQQLLFVQSF 90  
 Db 62 REALTEVRRAFHFLKSGRVRAVYIGELAMSTIENLFNRVLDRIASAPVQ----- 113  
 QY 91 GIPVSSSKLLQFLDQAVANDPQTLQONIMDKNMALHVEVQHGKSGGQTFHSLLTAS 150  
 Db 114 -----RVVDQVALLPELVEE-----FAAAGRQRDVDLLAATAHALAKKE 155  
 QY 151 ---LPRRDSTEAPRKSPSEQPIG-----OGRLVGTQLRVLGPEDDLAG 193  
 Db 156 PLRPPAPDDGGVP-PEAGADQPSLNDNGVAPRLADPAQAAMAGSDVELLDQ----- 209  
 QY 194 MFLQIFPLSPDKWQS-----SSPRVALAQALQGLARVQSSPEVGITVR 243  
 Db 210 -LLEIFTEAETHLEALVGLADCARELPOVTDALQAL-----HTLKSAHMGITL- 261  
 QY 244 VIALATLILSPHGALVMSHRSFLA-----CPLLRLQCOYORCVPOPTGFSSFLKV 298  
 Db 262 PIAETATPLEK-----LVKEYSNLAFDLREARELHDABQLFRIGLEOYGAORPLNPI 315  
 QY 299 ---LLOMLQWLDSPGV-----EGGPLRAQLRMLASQASAGRRLSDVROGLRLA 344  
 Db 316 PGSDALLERIEALHGERIASLEAERYSDAGERRPDLLEAFVBCMDILLDAEDLLERWH 375  
 QY 345 EALAFRODLEVSSTVAATATLRSQ-----ECSVEPDLISKVLOGLI----- 388  
 Db 376 EHPEROEL-----SALREELSTLDGRARHAEIPQVEELCAALLALAYAVEEGRLAVSPAF 431  
 QY 389 --EVRSPH-----LEELITA-----FFSATDAASAPP 414  
 Db 432 FEEARQAHREALIGMDVAAGLOVTPPERVAALQELLLEAPAAEAVFIPDESIGADDFP 491  
 QY 415 -----ACKPVVVVSSLLQEEEPPLAGKPGADGSL-----BAVRLGPSG-LLV 458  
 Db 492 PEDDEPALPEAVEEAGAPAEETVPAPAPAPGHEDEEMVSTFLEAVDILSAGALA 551  
 QY 459 DWLEMDPEVYSSCPDQLRLFLSRRKKGQAOVPSFRPYLLTLFHQSSMPTLHOCIRV 518  
 Db 552 QW-QAEPGALSSLSALQ-RDLHTLKGARMAEIAEIGD-----LAHE-----LEA 594  
 QY 519 LIGKSRQRPDPSASLDFLWACHIVPRIMOGROQT---PQKRREELVLRV---QGP-- 569

Db 595 IVEGLVDRHYHSPQLAGLQACH--DRLAEQDLQSLAGOPLADPHDLQIRFRROGPVA 653  
 QY 570 -----ELISIVELLIAEETRSDQDDTAACSLQIARPLLLSCCGDDESVKRYEHLS 623  
 Db 654 EAATPEASBPVEELVAPA---VEEPAAPAAEAEEERDEPELVEIFL---EEGFDILDSAA 707  
 QY 624 GCIOQWGSV----- 633  
 Db 708 AALQRMMDVDVNTIELEALQORLHTLKGAARMAEIGDGLAHELFYGLCGGRIRAS 767  
 QY 634 -----LGRRCRLLOLY-----LQR---PELRVPEVL---LH 662  
 Db 768 PALFGLQRCHELAEMLAVERGHRTLPDQALIAEIRLRSPDEQLSVPTSVSLKPLA 827  
 QY 663 SGAASSSVCKLDGLHRRFTLLAD-----TSSRALENGGADSMCRKIAVA 711  
 Db 828 AKGAADSEELIDFLEADQDLLENLELAGRMDSNGDQPLDD----- 872  
 QY 712 HPLLLRLPMIALHGRHNLNFOEFRQONHLSCEFLHYGLLELLOPHVFRSEHOGALN 771  
 Db 873 --LLRLHTLKGGARLAGQTELG-----NLAHDEQHLTDQOQCAPN 913  
 QY 772 -DCLSLFTLLNTRYKSSRHAAFINKEVQFIHKYTYNAPPAISFLQKHADPLHDSFD 830  
 Db 914 PDSL-----LDASGLEGLQORQVDLLRERLAE 941  
 QY 831 NSDLVWLKSLAG-----LSLPSRDDRTDR-----GLDEGESESSAGSLPLVSVSLFTFP 880  
 Db 942 DDE-----AGEPEPAQALVQAD--TDRAVASALAEITRLAPAGA--IMAAEAP 990  
 QY 881 LTAENAPMYKRLSRGQTEVLELVLSDIDEMSRRP-ELISFSTNLQRLMSAAE 936  
 Db 991 AAPATLTPVRKAQEAQ-----EASRRAPQELVKVPALLENLVNLAGE -1036

## RESULT 13

064720 PRELIMINARY; PRT; 1217 AA.  
 AC 064720;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE At2g02560 protein.  
 GN AT2G02560.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gfoll J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Cohenhaver G.P., Preuss D., Nierman J.C., White O., Eissen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.";  
 RT Nature 402:761-768(1999).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC004136; AAC18930.1; -;  
 SO SEQUENCE 1217 AA; 134621 MW; AA23A6A494EE1EA CRC64;

Query Match 2.8%; Score 149; DB 10; Length 1217;

Best Local Similarity 20.4%; Pred. No. 0.066;  
 Matches 196; Conservative 137; Mismatches 334; Indels 292; Caps 49;

QY 46 DTSEBALPLPDWIKLRIRRESEVLRINDAALQDEPQULLFYVSFGIPVSMKSLQFLD 105  
 Db 409 DTDESS---PKWL---LKEQSVKIVKINSIROREKSVKRVGAFSV---LRELVLVLP 457  
 QY 106 QAVAHDPQTL-----EONIMKNYMAHL-VEVQHERGASGGQTFSLTASLPREDSTEA 160  
 Db 458 DCLADHIGSLVPGIERLNDKSSISNLKIE-----ALVFYKVLASHAP----- 501  
 QY 161 P-----KPKSSPE-OPIGGRIRV-----GTOLRVLP----- 187  
 Db 502 PVHPHYIKALSSPVLAAGVGRYKVTAEALRVGCELYRVVRPSPAGKPFVHPITYN 561  
 QY 188 -----EDDLAGMFLQFPLSPBPMSSSPRPVLAALQCALQELARVY 231  
 Db 562 AIMSRLTNODQDEVEKCAITCMGLVSTFGDOLRAELPSCLPV---LVDRMGNEITRLT 618  
 QY 232 QGSPVEVGVIVRVLOALATLLSPFGALVSMHRSHPFLACPILRQLOQYORCPVODTGF 291  
 Db 619 -----AVKAFSVIATSP-----LH-----INLSCVIDHLIAELTGF 649  
 QY 292 SSLFLKYLQMLQWLDSPGVEGGPLRAQLRMLASQASA--GRRL--SDVRGGLRLAEALAF 349  
 Db 650 LRRANRVLRQ-----ATLITMNTLVATYGDKIGSEAYEIVLVELSSLIS 693  
 QY 350 RQLEVVSSIVRAVIATLRSGECCSVPEPLI--SKVL--QGLIEVRSHLEL---LTA 402  
 Db 694 VSDLHMHTALE--LCCGLMKGKSCSENI SLAVRKVYLPQALTLVKSPLQGLQALLDLOKF 752  
 QY 403 FSATA--DAASPFPACKFVVVVSSLL--LOEEEPFLAGRP-----GADGG 443  
 Db 753 FEALVYHANSF-----YTLESELSCAKPSQSGVPRKQALYSIAQVAVLCLAADKN 807  
 QY 444 SLEAVRLGPSSGLVDMLDEPVVSSCPDLQRLI---FSRRKG--KQAOVPSFRPY 498  
 Db 808 CSSTVKM-----LMEILKDSGTNSAKOHLALLSLGELGRKDLASHAGIET--- 854  
 QY 499 LTLFTHOSSMPRLHOCIRVLLGKSRQRPDSASLDELMACHIVPRIMQGRDRTFOKR 558  
 Db 855 -IYESFQSPPEETKSAASYALGNIAVGNL--SNYLPFI-----LDQIDNOOK 899  
 QY 559 REELVLRVQGPSELISVELLIAEATRSQDGDTAACSLQIARPLLLSCCGDDESVKRY 618  
 Db 900 KQYITLH-----SLKEVIVRQ---SVKADQNSVEKILALLNHCESEEGRVN 948  
 QY 619 TEHLSCGIOQWGSVIGRCRDLLQLYLQRPBELRVPEVLLHSEGAASSVC----- 672  
 Db 949 VAECLG-----KMALIEPEKLVAPALQVRTSPAAFTRAVTAVVKY 989  
 QY 673 -----KLDGL-----IHRFTTLADSDSRALENRQADSMACRKIAVAH--PLLLRH 719  
 Db 990 SYVERPEKIDELIIPQISSFLMLKLD--GDRHV--RRAVASALS---TFHAYKNLLKGL 1042  
 QY 720 LPMIALHGRHNLNFOEFRQONHLSCEFLHYL--GLELLLOPHVFRSEHOGALNDCLLSF 777  
 Db 1043 LPELPLLYQYIYK--KELIRYDLPKFKHYVDGL-----ELKRAAECEVFTL 1090  
 QY 778 IRLLYNRKSSRHAAFINKEVQ-----FIHKYTYNAPPAI--SEFLQKHADPLH 825  
 Db 1091 VDSCLQDVNPSSFIVPLFKSGLEDHYDKMLCHILSLADKCSAVLAVLDSLVERPLHK 1150  
 QY 826 DLSEF-----NSDLWLKSLLAGLSLPSRDDRTDRGIDEGEESAGSLPL 872  
 Db 1151 TISFKPKQADVKOEHRNED--MTRSLRAISSLDRLNGVYSHKFGGLMGCKMRKRSVPL 1207

## RESULT 14

09H7L6 PRELIMINARY; PRT; 1310 AA.  
 ID 09H7L6;  
 AC 09H7L6;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
 DE FLJ00056 protein (Fragment).  
 GN FLJ00056.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPLEEN;  
 RA Ohara O., Nagase T., Kikuno R., Okumura K.;  
 RT "The nucleotide sequence of a long cDNA clone isolated from human  
 spleen."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK024463; BAB15753.1;  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000219; RhoGEF.  
 DR Pfam: PF00169; PH.1.  
 DR SMART: SM00233; PH.1.  
 DR SMART: SM00325; RhoGEF.1.  
 DR PROSITE: PS50003; PH\_DOMAIN.1.  
 FT NON\_TER 1  
 SQ SEQUENCE 1310 AA; 141544 MW; E55E30759653C045 CRC64;

Query Match 2.88; Score 145; DB 4; Length 1310;  
 Best Local Similarity 19.7%; Pred. No. 0.14;  
 Matches 190; Conservative 87; Mismatches 284; Indels 402; Gaps 41;

QY 12 LTTGPPRADSEFQALLDIFPEKPLPTAFIVDT--EALLLPDWLKMIRSVL 68  
 DB 340 LTTTPP-----CPPEPPPSKDTNTLTHLHSLRPD----- 373  
 QY 69 RLVDAALDLEPOQLLVFVSFGIPVSSMSKILQFLDQA-----VAHDPQLE-- 116  
 DB 374 -----LQTGLSVLLDLRQAPRLPALIPALSQLDSDGPLVQRLLILHDLPLELC 427  
 QY 117 -----QNTMDKNYM-----AHLYVOHE-----RGASGC--Q 141  
 DB 428 GFOGAELVSENDLKRVAKPELQWELGHRDPSHWEIHOEVVRLCRLCQVIGSVRQ 487  
 QY 142 TFHSLTASLPRRDSTEARPKSSPEOPIGQGRIVGTQRLRVGPEDDLQAGMQLQIFPL 201  
 DB 488 AIEELGAAEEEEEAAVGMKPLQ-----KVLAPRLALQRD--GGAILMLRLS 535  
 QY 202 SPDRWSSSPRYVALALQALGQELARVYOGSPEVPGITVRVLAATLTLSSPHGALV 261  
 DB 536 TPSSKLEGGPATLYQEVDEAIHQ--LVR-----LSNLH----- 567  
 QY 262 MSMHRSHFLACPLRLQCYORCPQOTGFSSLFKVLQMLQWLDSPVE-----G 313  
 DB 568 -----VQOQEQRQC-----LRLQOVLQWLSGPEGQLASFPMPG 602  
 QY 314 GPLRA-----OLRMLASOASGRRLSDVR----- 337  
 DB 603 DTLALQETELRFAFSAEVOERLAQAREALALEENATSOQVLDIFEORLEQVSGILHRA 662  
 QY 338 -----GGILRLA-----EALAFRODLEVVSST--VRAVI----- 364  
 DB 663 LRLQRFQQAHEWVDEGFARLAGAGPREAVLALALRRAPESAGFQEKRALALDLGS 722  
 QY 365 -ATLRSEQCSVE--POLISKVLOGLIEVRSPHEELLTAFFSATADA-----SP 412  
 DB 723 PAALREMGRCQARCOELERRIOQHIGEEASPR-----GYRRRRADGASSGGAQWGPSP 776  
 QY 413 FPACKPVVVVSVLLQEEEPFLAGKPGADGSL-----EAVRLGPS-----GL 457  
 DB 777 SPS-----LSLLLPSS--PGRRPAPSHCSLAPCEDEEESPELAPEAGEPPRAVL 827  
 QY 458 VDWLEMLDPEVYSSCPDLQRLLFSSRRKGGQAQVPSFRPYLLTLFTHQSSMPTLHOCIR 517  
 DB 828 IRGLEVTSTEVVD-----RNCSPREH----- 848

QY 518 VLKSRQRPDPASDIFLMACIHPRIWQGRDQRTPOKREELVLRQ----- 567  
 DB 849 VLLGRAR-----GPDGPMG--VGTPMERKRSISAQORLVSELIACEQDYVATLSEPV 899  
 QY 568 -----GPELISVLELLAEFTRSQGDPTACSLIQAAL-----PLLS 606  
 DB 900 PPPPEL-----TPELKGTWMAALSAARERLSPFRTHTFLRELQCATHTPL-- 944  
 QY 607 CCCGDDESVRKVTLEHSCGICQMDSVLGRRCRLDLLQYLQRELPAPVEVLHSEGA 666  
 DB 945 -----RIGACFLRHGDQF-----SLYAYVYHRRKLENGLAALSPSSKGS 984  
 QY 667 ASSSVCKLDG-----IHRFTLLADTSDSRALENRGADASMACKLVAHPLL---L 717  
 DB 985 MEAGPYLPRAQLQPLEQFLTRFYGRLLFE-----LIREAGPELSSRCRALGAVALRQEA 1039  
 QY 718 RHLPMIALHGRTHLNQEFROQNLSCFLHYGLLELQPHYFRSHQALWDCILSF 777  
 DB 1040 RGRDILAVEAVRGEIDLEKGGQLLHRDFTVIGCRKKCLR--HVFLEH-----LLLF 1091  
 QY 778 IRL 780  
 DB 1092 SKL 1094

RESULT 15  
 Q8TER5  
 ID OTHERS PRELIMINARY; PRT; 1546 AA.  
 AC OTHERS;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE FLJ00128 protein (Fragment).  
 GN FLJ00128.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPLEEN;  
 RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;  
 RT "The nucleotide sequence of a long cDNA clone isolated from human  
 spleen."  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK074057; BAB84883.1;  
 FT NON\_TER 1  
 SQ SEQUENCE 1546 AA; 167672 MW; 7CEFE00757462910 CRC64;

Query Match 2.88; Score 145; DB 4; Length 1546;  
 Best Local Similarity 19.7%; Pred. No. 0.18;  
 Matches 190; Conservative 87; Mismatches 284; Indels 402; Gaps 41;

QY 12 LTTGPPRADSEFQALLDIFPEKPLPTAFIVDT--EALLLPDWLKMIRSVL 68  
 DB 576 LTTTPP-----CPPEPPPSKDTNTLTHLHSLRPD----- 609  
 QY 69 RLVDAALDLEPOQLLVFVSFGIPVSSMSKILQFLDQA-----VAHDPQLE-- 116  
 DB 610 -----LQTGLSVLLDLRQAPRLPALIPALSQLDSDGPLVQRLLILHDLPLELC 663  
 QY 117 -----QNTMDKNYM-----AHLYVOHE-----RGASGC--Q 141  
 DB 664 GFOGAELVSENDLKRVAKPELQWELGHRDPSHWEIHOEVVRLCRLCQVIGSVRQ 723  
 QY 142 TFHSLTASLPRRDSTEARPKSSPEOPIGQGRIVGTQRLRVGPEDDLQAGMQLQIFPL 201  
 DB 724 AIEELGAAEEEEEAAVGMKPLQ-----KVLAPRLALQRD--GGAILMLRLS 771  
 QY 202 SPDRWSSSPRYVALALQALGQELARVYOGSPEVPGITVRVLAATLTLSSPHGALV 261  
 DB 772 TPSSKLEGGPATLYQEVDEAIHQ--LVR-----LSNLH----- 803

```
QY 262 MSMSHSHFLACPLRLQYORCVPODGTGSSFLTKVLQMLQMLDPSGVE-----G 313
DB 804 -----VQOQROROC-----LRRLQOVLQMLSGPGEOLASFAMPG 838
QY 314 GPLRA-----QLRMLASQASAGRRLSDVR----- 337
DB 839 DTLASLOETELFRFASAEVQERLAQAREALALEENATSQKVLDFEQRLQEVESGLHRA 898
QY 338 -----GGLRLA-----EALAFRODLEVSST---VRAVI----- 364
DB 899 LRLQRFPOQAHFWDEGFARLAGCPGPAVLAALALRRAPESAGTQEMRALALDLS 958
QY 365 -ATLSGECQSYE-PDLISKVLQGLIEVSPHLELLTAFESATADA-----SP 412
DB 959 PALREMGRCQARQCELEIRIQHLOEASPR-----GYRRRADGASSGGQWGPSP 1012
QY 413 FPACRPVVVSSLLQEEPLAGKPGADGSL-----EAVRLGPS-----GLL 457
DB 1013 SPS-----LSSILLPSS---PGPRPAPSHCSLAPCGEDYEERGPDLAPAEAGRPRAVL 1063
QY 458 VDMLEMLDPEVWSSCPDLQRLRLFSRRKKGQAQVPSFRPYLLTLFTHOSWPTLHQCTR 517
DB 1064 IRLQEVSTEVVD-----RTCSPREH----- 1084
QY 518 VLLGSRQRFDPDSASLDPLMACIHVPRIWGRDQRTPOKRREELVLRVQ----- 567
DB 1085 VLLGRAR-----GPDGPMG-VGTPRMERKRSTISAOQRLVSELACQODYATLSEPV 1135
QY 568 ---GPELISLVELILAEFTRSODGPTACSLQARL-----PLLS 606
DB 1136 PPPGPRL-----TPELRGTWAAALASRELRSEFHTHFLRELOGCATHPV--- 1180
QY 607 CCCGDESVKVTENHLSGCIQOMGDSVLGRRCRDILLQYLQRPRLRVPVPEVLHSEGA 666
DB 1181 -----RIGACFLRHGDOF-----SLYAQYVKHRHKLNGIALSPSSKGS 1220
QY 667 ASSSVCKLIDGL-----IHRFTLLADTSDSRALENRGADASMACKRLVAHPILL---L 717
DB 1221 MEAGPYLPALQOPLQRLRYGRLEE-----LLREAGPELSEGCRAIGAAVQLREQEA 1275
QY 718 RHLPMIAALHGRTHLNRFQEFROQNHLSGFLVHVLGELLELOPHVERSEHOGALMDCLSF 777
DB 1276 RGRDLAAVEAVRGCEIDLKEGQQLHRDPFTYICGRKKCLR-HVFLFEH-----LLLF 1327
QY 778 IRL 780
DB 1328 SKL 1330
```

Search completed: May 7, 2003, 19:42:18  
Job time : 62 secs

**THIS PAGE BLANK (USPTO)**

---

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 13, 2003, 22:56:01 ; Search time 6720 Seconds  
(without alignments) 4456.367 Million cell updates/sec

Title: US-09-929-769-7  
Perfect score: 5241  
Sequence: 1 MHILVHAMVILLTLGPRA.....MDPSAQISEALRIHVEAVM 1029

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n model -DEV=x1h  
-O=/cogn2\_1/USPRO/spool/US09929769/runat\_07052003\_161924\_8034/app.query.fasta.1.1223  
-DB=GenEmbl -OPMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09929769 @CGN.1.1.4449 @runat.07052003.161924.8034 -NCPU=6 -ICPU=3  
-MARK=NO -MMAP -LARGEOUVR -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-MARK.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELPEXT=7

Database : GenEmbl : \*  
1: gb\_ba : \*  
2: gb\_htg : \*  
3: gb\_in : \*  
4: gb\_om : \*  
5: gb\_ov : \*  
6: gb\_pat : \*  
7: gb\_ph : \*  
8: gb\_pl : \*  
9: gb\_pr : \*  
10: gb\_ro : \*  
11: gb\_sts : \*  
12: gb\_sy : \*  
13: gb\_un : \*  
14: gb\_vl : \*  
15: em\_ba : \*  
16: em\_fun : \*  
17: em\_hum : \*  
18: em\_in : \*  
19: em\_mu : \*  
20: em\_om : \*  
21: em\_or : \*  
22: em\_ov : \*  
23: em\_pat : \*  
24: em\_ph : \*  
25: em\_pl : \*  
26: em\_ro : \*  
27: em\_sts : \*  
28: em\_un : \*

29: em\_vi : \*  
30: em\_htg\_hum : \*  
31: em\_htg\_inv : \*  
32: em\_htg\_other : \*  
33: em\_htg\_mus : \*  
34: em\_htg\_pln : \*  
35: em\_htg\_rnd : \*  
36: em\_htg\_mam : \*  
37: em\_htg\_vrt : \*  
38: em\_sy : \*  
39: em\_htgo\_hum : \*  
40: em\_htgo\_mus : \*  
41: em\_htgo\_other : \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5241	100.0	3437	6	AX468679 Sequence
2	5237	99.9	4434	9	AB037861 Homo sapi
3	4933.5	94.1	4179	9	BC004286 Homo sapi
4	4118	78.6	2769	9	BC013367 Homo sapi
5	3554	67.8	2418	9	BC018777 Homo sapi
6	3442.5	65.7	2730	10	BC010333 Homo sapi
7	3244	61.9	2241	9	HSMB00197 Homo sapi
8	3199.5	61.0	14896	2	AC102953 Homo sapi
9	2673.5	51.0	21964	2	AC130221 Mus muscu
10	1359	25.9	1141	9	HSMB002043 Homo sapi
11	865	16.5	157851	2	AC020509 Drosophi
12	863	16.5	146432	3	AC004642 Drosophi
13	863	16.5	154840	3	AC009018 Drosophi
14	863	16.5	303823	3	AE003462 Drosophi
15	738.5	14.1	144097	2	AC120290 Rattus no
16	431	8.2	144097	2	AC120290 Rattus no
17	374.5	7.1	6413	3	DMU24676 Drosophi
18	186	3.5	22693	3	HSPELE133 Homo sapi
19	175	3.3	165177	2	AC116507 Mus muscu
20	171.5	3.3	39896	1	SCK13
21	167	3.2	210017	2	AC102503 Mus muscu
22	164	3.1	4477	9	AK024475 Homo sapi
23	163.5	3.1	14800	6	AX334119 Sequence
24	163.5	3.1	14800	9	HSU53204 Human plect
25	163.5	3.1	189050	1	AL646077 Ralstonia
26	162	3.1	14189	9	HSPELECTIN
27	160.5	3.1	110000	2	LMFLCHR34_01
28	160	3.1	90348	1	AF497482
29	159.5	3.0	13946	1	AE004537
30	159	3.0	172148	3	LMFP656
31	157.5	3.0	15580	1	AE004478 Pseudomon
32	157	3.0	33529	6	AR166425 Sequence
33	156.5	3.0	5997	4	AF212149 Oryctolag
34	156.5	3.0	6038	4	AF212148 Oryctolag
35	156.5	3.0	6069	4	AF212147 Oryctolag
36	156.5	3.0	49377	6	188042
37	156.5	3.0	67523	1	SCU24241
38	156	3.0	7425	1	AX024365 Sequence
39	156	3.0	7425	6	AX024258 Sequence
40	156	3.0	47713	1	AX024320 Sequence
41	156	3.0	47713	6	AX024213 Sequence
42	155.5	3.0	11904	3	PAU79580 Pseudomonas
43	155	3.0	31252	3	LMFL6293
44	155	3.0	100185	8	AC114899
45	155	3.0	100957	8	AC116560

RESULT 1

## ALIGNMENTS

AX468679	LOCUS	AX468679	3437 bp	DNA	linear	PAT 16-JUL-2002
	DEFINITION	Sequence 3 from Patent WO0216581.				
	ACCESSION	AX468679				
	VERSION	AX468679.1	GI:21901457			
	KEYWORDS					
	SOURCE	human.				
	ORGANISM	Homo sapiens				
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	AUTHORS	1				
	TITLE	Gao,W.O., Polakis,P., Shou,J., Smith,V., Soriano,R., Williams,P.M., Wu,T.D. and Zhang,Z.				
	JOURNAL	Compositions and methods for the diagnosis and treatment of tumor Patent: WO 0216581-A 3 28-FEB-2002;				
	FEATURES	Genentech, Inc. (US)				
	source	location/Qualifiers				
		1..3437				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
BASE COUNT		558 a	1186 c	1098 g	595 t	
ORIGIN						

Alignment Scores:	
Pred. No.:	3,16e-315
Score:	5241.00
Percent Similarity:	100.00%
Best local Similarity:	100.00%
Query Match:	100.00%
DB:	6
US-09-929-769-7 (1-1029) x AX468679 (1-3437)	
	Length: 3437
	Matches: 1029
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Gaps: 0

QY	1	MethisIleValValIHisAlaMetValIleuLeuThrLeuGlyProProAla	20
Db	46	ATGCACATCCCTCGTGGTCATGCCATGTATCTCTGCACGCTGGGCGCCCTCGAGCC	106
QY	21	AspAspSerGluPheGlnAlaIleuLeuAspIleTrpPheProGluGlyAsProLeuPro	40
Db	106	GACGACACCGAGTTCACAGGCGCTCTGGACATCTGTGTTCCGGAGGAGAAAGCCATCGCCC	165
QY	41	ThrAlaPheLeuValAspThrSerGluGluAlaLeuLeuLeuProAspTrpLeuLysLeu	60
Db	166	ACCCCTCTTCGGTGGACACATCGAGAGAGGCGGTGCTCTCTGACTGCTGTAACCTG	225
QY	61	ArgMetIleArgSerGluValLeuArgLeuValAspAlaIleGluGlnAspLeuGluPro	80
Db	226	CGCATGATCCCTTCTGAGGTCTCCGCGCTGGTGGACGCCGCCCTCTGCAGACTGSAACCG	285
QY	81	GlnGlnLeuLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerLysLeu	100
Db	286	CAGCAGCTGCTGCTGTTCGTGCATCGTGGATCCCGGTCGACAGATGACCAATCTC	345
QY	101	LeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGluGlnAsnIleMet	120
Db	346	CTCCAGTCTCGGACGACGAGGTGGCCACACCCACAGATCTCGAGACGAGAACATCATG	405
QY	121	AspLysAsnTrpMetAlaHisLeuValGluValGlnHisGluArgGlyAlaSerGlyGly	140
Db	406	GACACAGATTTACATGCGCCACCTGTGTGAGAGTCCAGCATGACGCCGCGCTCCGGAGGC	465
QY	141	GlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgArgAspSerThrGlnAla	160
Db	466	CAGACTTCCACTCTCTTGCTCACAGGCTCCCTGGCGGCCGCGGACAGACAGAGAGCA	525
QY	161	ProLysProLysSerSerProGluGlnProIleGlyGlnGlyArgIleArgValGlyThr	180
Db	526	CCCAACACCAAGACAGCCACGACAGACCCCATTAAGCCAGAGGCCGAGATTCCGGTGGGAGCC	585
QY	181	GlnLeuArgValIleuGlyProGluLysAspLeuAlaGlyMetPheLeuGlnIlePhePro	200
Db	586	CAGTCTCGGGGTGGGCGCTCGAGAGACAGCTGGTGGCATGTTCTTCACGATTTTCGCG	645

QY	201	LeSerProAapProArgrTrpGlnSerSerSerProArgrProValAlaLeuAlaLeuGln	220
Db	646	CTCACCCCGGAGACCTCGGGGTGGACAGGCTCCAGTCCCGCCCGGGGCTCGGCCCTGGAC	705
QY	221	GlnAlaLeuGlnGlnGlnLeuAlaArgValValGlnGlnSerProGluValProGlnIle	240
Db	706	CAGGCCCTGGGGCAGAGACTGGCCGGGTGTGTCCAGGGCAAGCCCGAGGTGCCGGGATC	765
QY	241	ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyGlyAlaLeu	260
Db	766	ACGGGTGCTGTCTGGACAGCCCTTGCACACCTGTCTCAAGTCCACACGGCGGTGCCCTG	825
QY	261	ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuCysGln	280
Db	826	GTGAGTGTCCATGGACACCGTAGCCACTTCTGTGGCCCGCGGTGGCCGACGTCTGCAG	885
QY	281	TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValLeuLeu	300
Db	886	TACACGCCCTGTGTGCCACAGAGACCGGGCTTCTCTCGCTTCTCCGAAGGTGCTCTG	945
QY	301	GlnMetLeuGlnTrpLeuAspSerProGlyValAlaGlnGlyProLeuArgAlaGlnLeu	320
Db	946	CAGATGCTGCAGTGGCTGTGACAGCCCTGTGGCTGGAGGGCCGGCCCTCTGGCGGCACACT	100
QY	321	ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeu	340
Db	1006	AGGATGCTTCCACACAGGGCTTCCAGCGGGCGAGCCTCATGTGCTCCAGCGGGGGCTC	1066
QY	341	LeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValValSerSerThrVal	360
Db	1066	CTGGCCCTGGCCCGAGGCCCTGTGGCTTCCGTACAGACTGTGAAGGTGTGTACTCTCACCTC	112
QY	361	ArgAlaValIleAlaThrLeuArgSerGlyGlnGlnCysSerValGlnProAspLeuIle	380
Db	1126	CGTCCCGCATTCGCGACACCTTAGGTCTGGGAGAGCAGTGCAGCGTGGAGCGGACCTGATC	118
QY	381	SerIysValLeuGlnGlnGlyLeuIleGlnValArgSerProHisIleGlnGlnIleLeuThr	400
Db	1186	AGCAAAAGTCTCCAGAGGGGTGATGAGCTGAGTCCGCCACCTGTGAGGAGCTGTGACT	124
QY	401	AlaPhePheSerAlaThrAlaAspAlaAlaSerProPheProAlaCysLysProValVal	420
Db	1246	GCATTCCTCTGTGGACATGGGAGATGTCGTCGCCCGGTTTCAGACTGTAAAGCCGTTTG	130
QY	421	ValValSerSerLeuLeuLeuGlnGlnGlnGlnProLeuAlaGlyGlyLysProGlyAla	440
Db	1306	GTGGTGACTCTCTGTGTCTGCACAGAGAGAGAGCCCTGTGCTGGGGGGAAGCCGGGTGG	136
QY	441	AspGlySerLeuGlnAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrp	460
Db	1366	GACGTGGCACACTGTGAGCCGTGGCTGTGGGCCCTGTCAAGGCTCCTAAGTGCACGTGG	142
QY	461	LeuGlnMetLeuAspProGlnValValValSerSerCysProAspLeuGlnLeuArgLeuLeu	480
Db	1426	CTGGAATAAGTGTGGACCCCGAGGTGTGCAGAGCTGCCCCACACTGCAGCTGCAGGCTGTG	148
QY	481	PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProIyTrpLeuLeu	500
Db	1486	TTTCTCCCGAGAGAGGCAAAAGTCAAGCCAGGTGCTGTTCCTGTTCCTCACTCTCTG	154
QY	501	ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeuLeu	520
Db	1546	ACCCTCTCACGATCAGTCACTGAGCTGTGCCACACTGCACAGTGCATCCGAGCTCTGTG	160
QY	521	GlyLysSerArgGlnGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys	540
Db	1606	GGCAGAGACCCGGGAACAGAGTTTGACCCCTGTGCTCTCTGTGACATTCCTCTGTGGCTGTG	166
QY	541	IleHisValProArgIleTrpGlnGlyArgAspGlnArgThrProGlnLysArgArgGln	560
Db	1666	ATTCATGTCTCTCCCATCTGGCAGAGGGCGGGACACAGGCACCCCGCAGAAAGCGCGGAG	172
QY	561	GlnLeuValLeuArgValGlnGlyProGlnIleuIleSerLeuValGlnLeuIleAla	580



ELDQAVADPOTLEQINIMDKNYMAHLVEVOHERGASGQTFHSLTASLPERRDSTEA  
 PKRSPEDQPIGGRVCTOLRVLEPEDLAGMFOIQLPLSDPQWSSSPRPVLA  
 LQALGDELARVQSGEPTGIVRLVALATLMLSPHGLAWHNRHSHFLACPLLR  
 OCGORVOPDPMGFSSFLKVLQMLQWLDSPGVGGLRQVRLMLAQAQASRLS  
 DVGGLRLAEALAPRODELVYSTVYANVLTIRSEOCSEVEDRLSKYLOGLIEVRS  
 PHUBELITTFESATADASPFPACKPVVYSSLLDEEERLGGKAGSGSLAEVRL  
 GSSSGLVLMLEMLDEEVSSCPDLOLRLFSNRKKGQADVPFRPYLTLETHQSS  
 WPTLHQCIVLLKLSREORFDEPSADLRLMACIHYRIRWQROPTQKRELVLYRV  
 OGPBELSLVELLAEATRSODPTMACSLQIARPLLLSCCGDDESVKRYETHLG  
 CIOQMGDSVLRGRCDLQLOLREPLVPEVYELHSEGAASSVCKLDGLIHERFI  
 TLADTSDRALENRGDASMACRLAVAPILLRLHLMIALALHGRTHLNAPFERRO  
 QNHLSLFLHVLGLLELLOPHVRESEHQGLMPCLSITLLNLRKSSHRLAIFIKF  
 VQPIHRTYTNPAATISLQKADPDLHDSFQSDSLVMLKSLAGLSLSDRDRTRG  
 LDEESESASGLPLVSVSLPTPLAAMAPYMKRLSGQVLEDEVLSDIDEMSR  
 RRPELISFSTNQLRMSAECCRMALASLALRSMNSPISIAALPTPMYCLGSD  
 FEVQYATLNEPEYALICEHAAVLHRAFLVGMYGOMPSAQISALRLHMEAVY"

BASE COUNT 731 a 1521 c 1419 g 763 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 7,91e-315 Length: 4434  
 Score: 5237.00 Matches: 1028  
 Percent Similarity: 99.90% Conservative: 0  
 Best Local Similarity: 99.90% Mismatches: 1  
 Query Match: 99.92% Indels: 0  
 DB: 9 Gaps: 0

US-09-929-769-7 (1-1029) x AB037861 (1-4434)

QY 1 MeHHisIleuValValHisAlaMetValIleleuLeuThrIleuGlyProProArgAla 20  
 Db 1045 ATGACATCTCTCGTGGCCATGCCATGCTGTCAGCGGCGCCGCTCGAGCC 1104  
 QY 21 AspAspSerGluPheGlnAlaLeuLeuAspIleTrpPheProGluGluProLeuPro 40  
 Db 1105 GACGACAGCAGTCCAGCGCGCTGTCGACATCGGTTCCGGAGGAGAACCCCTGCC 1164  
 QY 41 ThrAlaPheLeuValAspThrSerGluGluAlaLeuLeuProAspTrpLeuLysLeu 60  
 Db 1165 ACCGCTCTCTGGTGGACATCGAGAGCGCGCTGCTTCTGCTGCTGAACTG 1224  
 QY 61 ArgMetIleArgSerGluValIleuArgLeuValAspAlaAlaLeuGlnAspLeuGluPro 80  
 Db 1225 CGCATGATCCGTTCTAGAGTGCTCCGCTGCTGAGCCGCCCTCGCAGAGCATGAGACCG 1284  
 QY 81 GlnGlnLeuLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerLysLeu 100  
 Db 1285 CACGACGTCCTGCTGTCGTCAGTCTGTTGGCATCCCGTGTCCACATGAGCAAACTC 1344  
 QY 101 LeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGluGlnAsnIleMet 120  
 Db 1345 CTCACAGTCTCTGGACGAGGAGTGGCCACGACCCCACTCGAGAGCAAAATCATG 1404  
 QY 121 AspLysAsnTrpMetAlaHisLeuValGlnIleHisGlnArgGlyAlaSerGlyGly 140  
 Db 1405 GAAAGAAATTCATGCCCCACCTGCTGAGGTCACGATAGCGCGCCCTCGGAGGC 1464  
 QY 141 GlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgArgAspSerThrGluAla 160  
 Db 1465 CACACTTTCACACTCTTGTGCACAGCTCCCTGCGCCCGCGAGACAGACAGAGCA 1524  
 QY 161 ProLysProLysSerSerProGluGlnProIleGlyGlnGlyArgIleArgValGlyThr 180  
 Db 1525 CCCAAACCAAGAGACCCAGAGCAGCCATAGGCGGCGGATTCGGGTGGGACC 1584  
 QY 181 GlnLeuArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeuGlnIlePhePro 200  
 Db 1585 CAGCTCGGGTGGTGGGCGCTGAGAGCAGCTGGCTGGCATGTTCCTCGCAATTTTCCG 1644  
 QY 201 LeuSerProAspProArgTrpGlnSerSerSerProArgProValAlaLeuAlaLeuGln 220  
 Db 1645 CTGAGCCCGGAGCCCTGGTGGCAGAGCTCCAGTCCCGCCCGCTGCGCTCGAG 1704

QY 221 GlnAlaLeuGlnGlnGlnLeuAlaArgValValGlnGlySerProGluValProGlyIle 240  
 Db 1705 CAGGCGCTGGGCGAGGAGCTGGCGCGGCTGCTCAGAGGAGCCCGAGGTGCCGGCATC 1764  
 QY 241 ThrValArgValLeuGlnAlaLeuAlaThrLeuLeuSerSerProHisGlyAlaLeu 260  
 Db 1765 ACGTGGTCTCTGGACGCGCTCGCCACCTGCTGAGTCCACACGCGGCTGCCCTG 1824  
 QY 261 ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnIleCysGln 280  
 Db 1825 GTGATGCTTCATGACAGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1884  
 QY 281 TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuLysValLeuLeu 300  
 Db 1885 TACACGCGCTGTGGCCACAGCAGCACCGGCTTCTCCCTGCTTCCCTGAGAGGTCTCTG 1944  
 QY 301 GlnMetLeuGlnIleThrLeuAspSerProGluValGlnGlyLysProLeuArgAlaGlnLeu 320  
 Db 1945 CAGATGCTGAGTGGTGGACAGCAGCTGCTGAGGAGGCGGCGCGCTGCGGACAGCTC 2004  
 QY 321 ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeu 340  
 Db 2005 AGATGCTTGCAGAGCAGGCTCAGCGCGGCGAGGCTCAGTATGTCGAGGGGGCTC 2064  
 QY 341 LeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValValSerSerThrVal 360  
 Db 2065 CTGGCGCTGGCGCGAGCGCTGCGCTTCTGTCAGAGCTGAGAGGTGGTCACTCCACCTC 2124  
 QY 361 ArgAlaValIleAlaThrLeuArgSerGlyGluGlnCysSerValGluProAspLeuIle 380  
 Db 2125 CGTCCGCTCATCGCCACCTGAGGTCTGGGAGCAGTGCAGCTGGAGCGGAGCTGATC 2184  
 QY 381 SerLysValLeuGlnGlnGlyLeuIleGluValArgSerProHisLeuGlnGluLeuThr 400  
 Db 2185 AGCAAACTCTCCAGGAGGCTGATGAGTGAAGTGCACCCACCTGAGAGAGCTGCTACT 2244  
 QY 401 AlaPheSerSerAlaThrAlaAspAlaAlaSerProPheProAlaCysLysProValVal 420  
 Db 2245 GCATCTTCTCTGCGCATGGGAGTGGCTCCCTCCCGTTTCCAGCTGTAAAGCCGTTGTG 2304  
 QY 421 ValValSerSerLeuLeuLeuGlnGlnGluGluGluProLeuAlaGlyLysProGlyAla 440  
 Db 2305 GTGTGATGCTCCCTGCTGCTGACAGAGAGAGAGGCCCTGCTGGGGGAGAGCCGGTGC 2364  
 QY 441 AspGlyLysSerLeuGlnAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrp 460  
 Db 2365 GACGCTGACAGCTCGAGGCGCTGCGGCTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTG 2424  
 QY 461 LeuGlnMetLeuAspProGluValValSerSerCysProAspLeuGlnIleArgLeuLeu 480  
 Db 2425 CTGAAATGCTGGACCCCGAGGTGGTCAAGCAGCTGCCCGCACTGACAGCTCAAGCTGCTC 2484  
 QY 481 PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProTrpLeuLeu 500  
 Db 2485 TTCTCCCGGAGGAAGGCAAGGTCAGAGCCAGGTGCTCTGCTTCCCTGCTTCCCTGCTC 2544  
 QY 501 ThrLeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleArgValLeuLeu 520  
 Db 2545 ACCCTTTCACGATCAGCTGACGCTGCGCACACTGACACAGATGACATCCACAGTCTGCTG 2604  
 QY 521 GlyLysSerArgGlnGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCys 540  
 Db 2605 GGCAGAGCCGGGAAACAGAGATTCGACCCCTGCTTCTGACCTTCTGCTGCTGCTGCTG 2664  
 QY 541 IleHisValProArgIleTrpGlnGlyArgAspGlnArgThrProGlnLysArgArgGlu 560  
 Db 2665 ATCCATTTCTCTGCACTCTGCGAGGGCGGGGAGCAGCGCACCCGCGCAAGAGCGGGAG 2724  
 QY 561 GlnLeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGlnLeuIleLeuAla 580  
 Db 2725 GAGCTGCTGCTGGGCGTCCAGAGGGCGGAGCTCATGAGCTGAGTGCATGCTGCTGCTG 2784  
 QY 581 GlnAlaGlnThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnAlaArg 600

Db	2785	GAGCGGAGAGCGGGAGCCGAGGAGCGGGAGACAGCGCGCTCGAGGCTCATCCAGGCCGG	2844
QY	601	LeuProLeuLeuLeuSerCysCysGlyAspAspGluSerValArgLysValThrGlu	620
Db	2845	CTGGCCCTGGCTGCTCACTCTCTCTCTGGGGAGAGTAGAGTGTAGAGAGGTGACGAG	2904
QY	621	HisLeuSerGlyCysIleGlnGlnIleTPGlyAspSerValLeuGlyVARGArgCysArgAsp	640
Db	2905	CACCTGTACAGGCTGTCACTCAACAGCTGGGGAGAGAGCGTCTGTGGAGGGGGTGGCGAAG	2964
QY	641	LeuLeuLeuGlnLeuTyrlLeuGlnArgProGluLeuArgValProValProGluValLeu	660
Db	2965	CTTCTCTCGAGCTCTACAGCGCGAGAGCTCGGGGTGCCCGTGGCTGTGATCTCTTA	3024
QY	661	LeuHisSerGlnGlyAlaHisSerSerValCysLysLeuAspGlyLeuIleHisArg	680
Db	3025	CTGGACACGGCAAGGGCTGTCCACACAGCGTGTGCMACTGGAGCGACTCTCTCACCGC	3084
QY	681	PhelIleThrLeuLeuAlaAspThrSerAspSerArgAlaLeuGlnAsnArgGlyAlaAsp	700
Db	3085	TTTCATCAGCGCTCTTGGGGACACACAGAGCTCCCGGGCGTTGGAGACCGAGGGCGGAT	3144
QY	701	AlaSerMetAlaCysArgLysLeuAlaValAlaHisProLeuLeuLeuArgHisLeu	720
Db	3145	GGCAGCATGGCTCTGGGAAGCTGGGGGGGGCGCCACTCTCTCTCTGTCTGAGACCTG	3204
QY	721	ProMetIleAlaAlaLeuLeuHisGlyArgThrHisLeuAsnPhenGlnIleArgGln	740
Db	3205	CCCATGTATCGGGCGCTCTCTGACAGGGCGCCACCCACTTAACCTTCAGAGAGTTTCCGGCAG	3264
QY	741	GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuProHis	760
Db	3265	CAGAACCACTGAGCTCTCTCTGACAGCTGTGGGCTCTGGAGCTGTCTGCACGCCGAC	3324
QY	761	ValPheArgSerGlnHisGlnGlyAlaLeuTPAspCysLeuLeuLeuSerPheIleArgLeu	780
Db	3325	GTTGTCGGCAGCGGACCCAGCGGGGGGGCTGTGGAGCTGCTTCTGTCTCATCCGCTG	3384
QY	781	LeuLeuAsnTyrlArgLysSerSerArgHisLeuAlaAlaPheIleAsnLysPheValGln	800
Db	3385	CTGCTGAATTATACAGGAAGCTCTCCCGCATCTGGCTGCTTCATCAACAAGTTGTGGAG	3444
QY	801	PhelIleHisLysTyrlIleThrTyrlAsnAlaProAlaAlaIleSerPheLeuGlnLysHis	820
Db	3445	TTTCATCATAGATACATTACTCTCAATATGCCACGACCATCTCTCTCTGCAGAGACAC	3504
QY	821	AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu	840
Db	3505	GGCAGACCGCTCCACGACACTGTCTCTTCGACAAAGAGACTGTATGTGAATCCCTC	3564
QY	841	LeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArgGlyLeuAspGlnGlu	860
Db	3565	CTTTCACAGGCTCAGCCCTGCCACAGGAGGACAGAGAGCCAGCGAGCTGTGCACAAAG	3624
QY	861	GlyGlnGlnGlnSerSerAlaGlySerLeuProLeuValSerValSerLeuPheThrPro	880
Db	3625	GGCAGAGAGAGACTATAGCCGGCTCTTGGTCCCTTGTAAGCGTCTCCCTGTTCACCCCT	3684
QY	881	LeuThrAlaAlaGlnMetAlaProTyrlMetLysArgLeuSerArgGlyGlnThrValGln	900
Db	3685	CTGACCCCGCGCGAGTGGGCCCTCTCAATGAACGGCTTCCCGGGGCCAAACGCTGGAG	3744
QY	901	AspLeuLeuGlnValLeuSerAspIleAspGlnMetSerArgArgArgProGluIleLeu	920
Db	3745	GATTTGCTGGAGGTTCTGAGTGTACATAGAGAGATATCTCCGGCGGAGACCCGCAATCTGTG	3804
QY	921	SerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGlnLysCysArgAsn	940
Db	3805	AGCTTCTTTCGACCAACTGTCAGCGGGCTATAGCTCGCGCCGAGAGAGTGTGCCGACAC	3864
QY	941	LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaPhe	960

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	JOURNAL	REMARK	COMMENT
Db	3865	CTGCCTTCACGCTCGGCGCCCTCGCTCCATGACGAACACGCCACGATTCGACGCCCTTC	3924								
Qy	961	LeuProthrpPheMetTyrCysLeuGlySerGlnsppHegluValValGlnThrAlaLeu	980								
Db	3925	CTGGCCACGTTCCATGACTGCTGCTGCGGACGAGCCTTTGAGTGTGCAACAGCCCTTC	3984								
Qy	981	ArgAsnLeuProGluTyrAlaLeuLeuGlyGlnGluHisAlaValLeuLeuHisArg	1000								
Db	3985	CGGAACCTGCTGAGTACGCTCTCTCTGTGCGCAAGACGACGGGCTGTCTGTCACCGG	4044								
Qy	1001	AlaPheLeuValGlyMetTyrGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeu	1020								
Db	4045	GCTTCCGTCGTGGCATGTACGGCCAGATGAGCCACCGCAGATCTCGAGGCGCTG	4104								
Qy	1021	ArgIleLeuHisMetGluAlaValMet	1029								
Db	4105	AGGATCCGTCATATGAGAGCCGTCATG	4131								
RESULT 3											
LOCUS	BC004286	4179 bp	mRNA	linear	PRI 12-JUL-2001						
DEFINITION	Homo sapiens, clone IMAGE:3618123, mRNA, partial cds.										
ACCESSION	BC004286										
VERSION	BC004286.1	GI:13279124									
KEYWORDS											
ORGANISM	Homo sapiens.										
SOURCE	Homo sapiens										
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;										
AUTHORS	Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.										
TITLE	1 (bases 1 to 4179)										
JOURNAL	Strausberg,R.										
REMARK	Direct Submission										
COMMENT	Submitted (01-MAR-2001) National Institutes of Health, Mammalian										
	Gene Collection (MGC), Cancer Genomics Office, National Cancer										
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,										
	USA										
	NIH-MGC Project URL: http://mgc.ncl.nih.gov										
	Contact: MGC help desk										
	Email: cgaps-remail.nih.gov										
	Tissue Procurement: ATCC/DC/DP										
	CNA Library Preparation: Rubin Laboratory										
	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)										
	DNA Sequencing by: Genome Sequence Centre,										
	BC Cancer Agency, Vancouver, BC, Canada										
	Info@cgsc.bc.ca										
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,										
	Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,										
	Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo										
	San Lee, Victor Ling, Carrie Mathewson, Candice McLevey, Steven										
	Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeei, Jacqueline										
	Scheul, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stout,										
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,										
	George Yang, Scott Zuyderduyn, Marco Marra.										
	Clone distribution: MGC clone distribution information can be found										
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov										
	Series: IRAL Plate: 13 Row: 3 Column: 11				</						

[illegible]

OY	241	ThValArgValLeuGlnAlaLeuAlaAlaThrLeuLeuSerProHisGlyValLeu	260
Db	1632	ACGGGCGGTGCTCGCAAGGCGCTGGCAACCTCTCAAGTCCCAACAGCGGCGTCCG	1691
OY	261	ValMetSerMetHisArgSerHisPheLeuAlaCysProLeuLeuArgGlnLeuCysGln	280
Db	1692	GTGATGTCCATGCAACCGTAGCAACTTCTGCGCTGCGCGGTGGTGGCAAGCTTGCCAG	1751
OY	281	TyrGlnArgCysValProGlnAspThrGlyPheSerSerLeuPheLeuValLeuLeu	300
Db	1752	TACAGCGCTGTGCGCAAGGACACCGGCTTCTCTCCGCTCTCTGAAAGGTGCTCG	1811
OY	301	GlnMetLeuGlnThrPheAspSerProGlyValGlnGlyProLeuArgAlaGlnLeu	320
Db	1812	CAGATGCTGCAAGTGGCTGGACAGCCCTGGCGTGGAGGCGGCGCCCTGGCGGCACAGCTC	1871
OY	321	ArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgLeuSerAspValArgGlyLeu	340
Db	1872	AGGATGCTTGCCACGACGAGGCTCGAGCGGCGCCAGGCTCAGTGAATGTGGAGGCGGCTC	1931
OY	341	LeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValValSerSerThrVal	360
Db	1932	CTGGCGCTGGCGGAGGCGCTGGCTTCTCGACAGCACTGGAAGTGGTCAAGCTCCACGCT-	1990
OY	361	ArgAlaValIleAlaThrLeuArgSerGlyGlnGlnCysSerValGlnProAspLeuIle	380
Db	1990	-----	1990
OY	381	SerLysValLeuGlnGlnIleGlnValArgSerProHisLeuGlnIleLeuLeuThr	400
Db	1990	-----	1990
OY	401	AlaPhePheSerAlaThrAlaAspAlaAlaSerProPheProAlaCysLysProValVal	420
Db	1991	-----CCGTTTCCAGCCGTGAAGCCGCTTGG	2017
OY	421	ValValSerSerLeuLeuLeuGlnGlnGlnIleProLeuAlaGlyGlyLysProGlyAla	440
Db	2018	GTGGGAGACTCCCTCGTCGTCGCAAGAGAGAGACCCCTGGCTGGGGGAAAGCGGGCTCG	2077
OY	441	AspGlySerLeuGlnAlaValArgLeuGlyProSerSerGlyLeuLeuValAspThr	460
Db	2078	GACGGTGGCACTCGGAGGCGCTGGCGCTGGGCGCTCGTCAAGGCTCTCTAGTGGACGTGG	2137
OY	461	LeuGlnMetLeuAspProGlnValValSerSerCysProAspLeuGlnLeuArgLeuLeu	480
Db	2138	CTGGAAATGCTGGGACCCCGAGAGTGGTACGAGCTGGCCCGAAGCTGACAGCTCAGGCTTGCTC	2197
OY	481	PheSerArgArgLysGlyLysGlyGlnAlaGlnValProSerPheArgProThrLeuLeu	500
Db	2198	TTTCTCCGGAGAGGCGCAAAAGTCAGAGGCCACAGTCCCTCGTGGCTCTCACTGACTCGTG	2257
OY	501	ThrLeuPheThrHisGlnSerSerThrProThrLeuHisGlnCysIleArgValLeuLeu	520
Db	2258	ACCTCTTTCACGATCCAGTCCAGTGGCCACACTGCACACAGTCAATCCGACTCTCTGCTG	2317
OY	521	GlyLysSerArgGlnArgPheAspProSerAlaSerLeuAspPheLeuThrPheLacys	540
Db	2318	GGCAGAGCGGAGACAGAGTTGACCCCTTGCTCTGTGGACTTCTCTGGGCGCTGGC	2377
OY	541	IleHisValProArgIleArgGlnGlyArgAspGlnArgThrProGlnLysArgGln	560
Db	2378	ATTCATGTCTCTGCACATCTGGCAAGGGCGGAGACAGCGACCCCGAGAGAGCGCGGAG	2437
OY	561	GluLeuValLeuArgValGlnGlyProGlnLeuIleSerLeuValGlnLeuIleLeuAla	580
Db	2438	GAGCTGTGTGCTGGCGGTCCAGGGCCCGGAGCTCATAGCCTGTGGAGCTGATCTCTGGCC	2497
OY	581	GluLysIleuThrArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIleGlnIleArg	600
Db	2498	GAGCGGAGAGCGGAGCCAGACGGGGAGACAGCCGCTGAGACCTCATTCAGGCGCCGG	2557



Oy	601	LeuProLeuLeuLeuSerCysCysGlyAspAspGluSerValArgylsValThrGlu	620
Db	2558	CTGGCCCTGGCTGCTCACTCTCTCTGTGGGAGAGATGTGCAGGAAGGTGCAGAG	2617
Oy	621	HisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArgGlyAsp	640
Db	2618	CACCTGTACAGGCTGCATCCACGAGTGGGGAGACAGCGTCTGGGACAGCGCTGCCGAGAC	2677
Oy	641	LeuLeuLeuGlnLeuTyrlleuGlnArgProGluLeuArgValProValProGluLeu	660
Db	2678	CTTCTCTGGAGCTCTACACAGCGGCGGAGACTCGGGGTGCCGCTGACTGAGGTCTCTA	2737
Oy	661	LeuHisSerGlnGlyAlaAlaAspSerSerValCysGlyLeuAspGlyLeuIleHisArg	680
Db	2738	CTGCACACGGAAAGGGGTGGCCACACAGCGTGTGCMACTGGACGGACTTCATCCACCGC	2797
Oy	681	PhelIleThrLeuLeuAlaAspTrpSerAspSerArgAlaLeuGlyAlaAsp	700
Db	2798	TTTCATCACGCTCTTGGGGACACCACGAGCTCCCGGGCTTGGAGACCGAGGGCGGAT	2855
Oy	701	AlaSerMetAlaCysArgGlySleuAlaValAlaHisProLeuLeuLeuLeuArgHisLeu	720
Db	2858	GCCAGCATGGCTCCCGGAGACTGGGGGTGGCCACCCCGCTGCTGTCTGACGCACCTG	2917
Oy	721	ProMetIleAlaAlaLeuLeuHisGlyArgTrpHisLeuAsnPhelGlnIleuPheArgGln	740
Db	2918	CCCATGTATCCGGGGCTCCGACAGCGCGCACCCACTTCAGAGATTTCCGGAGC	2977
Oy	741	GlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuGlnLeuLeuLeuProHis	760
Db	2978	CAGAACCACTGAGCTCTTCTGTCACAGTCTGTGGCTCTGTGAGCTGTGCAGCCGCAC	3037
Oy	761	ValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSerPheIleArgLeu	780
Db	3038	GTTGTTCCGACGACGACACAGGAGGGCGCTGTGGAGCTGCTTCTGCTTCATCCGCTG	3097
Oy	781	LeuLeuAsnTrpArgGlySerSerSerArgHisSleuAlaAlaPheIleAsnLysPheValGln	800
Db	3098	CTGTGTGAATTACAGAGAGTCTCCCGCACTGTGGCTTCATCAACAAGTTTCTGGAG	3157
Oy	801	PhelIleHisLysTrpIleThrTrpAsnAlaProAlaAlaIleSerPheLeuGlnLysHis	820
Db	3158	TTTCATCCATAGTACATTACTTACATGCCCCACGACCATCTCTTCTGTGCAGAGACAC	3217
Oy	821	AlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeu	840
Db	3218	GCCGACCGGCTCCACGACGTGCTCTTGACAAAGTACGCTGGTATGCTGAATCCCTC	3277
Oy	841	LeuAlaGlyLeuSerLeuProSerArgAspAspArgTrpHisAspArgGlyLeuAspGlnGlu	860
Db	3278	CTTGCAGAGGCTCACCTCCGCCACAGGAGGACAGAGACCGACCGAGCTGTGCACCAAG	3337
Oy	861	GlyGlnGlnGlnSerSerArgIlySerLeuProLeuValSerValSerLeuPheThrPro	880
Db	3338	GGCAGAGAGAGACTTACGCCGGCTCTTGCCCTGTGTAGGGCTTCCCTGTTCAACCTT	3397
Oy	881	LeuThrAlaAlaGlnMetAlaProTyMetLysArgLeuSerArgGlyGlnThrValGln	900
Db	3398	CTGACCGCGGCGGAGTGGGCCCTCATGTAAMAAGGCTTTCCGGGGCCAAAGCGGTGAG	3455
Oy	901	AspLeuLeuGlnValLeuSerAspIleAspGlnMetSerArgArgArgProGluIleLeu	920
Db	3458	GATCTGTCTGGAGTTCGTGAGTGCATATAGACGATGTCCCGCGGAGACCCGAGATCTCTG	3517
Oy	921	SerPhePheSerTrpAsnLeuGlnArgLeuMetSerSerAlaGlnGlnCysGlyAsnArgSn	940
Db	3518	AGCTTCTTCTGACCACTGCAGCGGCTATAGCTGGCGGCGAGAGAGTGGCCGACAC	3577
Oy	941	LeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIleAlaAlaAlaPhe	960
Db	3578	CTGGCCCTTACGCGTGGCGCTGCGGCTCATGTACAGAAACGCCCAAGCATTTGCAGCTTTC	3637
Oy	961	LeuProThrPheMetTrpCysLeuGlySerGlnAspPheGlnValGlnThrAlaLeu	980

Db	3638	CTGCCACGTCATGTACTGCTTGGCGACCCAGGACCTTGAGTGTCACAGCGCCTC	3697
OY	981	ArganLeuProGluTyTAlaLeuLeuLysGlnGlnHisAlaValLeuLeuHisArg	1000
Db	3698	CGGAACGCTGTAGTAGCGCTCTCTGTGGCAAGAAGCACGGCGTGTCTCTCACCGG	3757
OY	1001	AlaPheLeuValGlyMetTyRGlyGlnMetAspProSerAlaGlnIleLeuGlnAlaLeu	1020
Db	3758	GCTTCCTCGTGGGCATGTACGGCCAGATGAGACCACCGCGCATCTCGAGGCGCTG	3817
OY	1021	ArgIleLeuHisMetGlnAlaValMet	1029
Db	3818	AGGATCTGCATATGAGACCGCGTATG	3844
RESULT 4			
LOCUS	BC013367	2769 bp	mRNA linear PRI 04-SEP-2001
DEFINITION	Homo sapiens, clone IMAGE:3677373, mRNA, partial cds.		
ACCESSION	BC013367		
VERSION	BC013367.1	GI:15426515	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2769) Strausberg,R. Direct Submission Submitted (31-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA NIH-MGC Project URL: http://mgc.ncl.nih.gov Contact: MGC help desk Email: gcgabs-remail.nih.gov Tissue Procurement: Louis M. Staudt, M.D., Ph.D. CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Institute for Systems Biology http://www.systemsbio.org contact: amadansystemsbiology.org Anup Madan, Rachel Dickhoff, Jessica Fahy, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan		
REMARK			
COMMENT			
FEATURES			
SOURCE			
CDS			







Db	361	CCCTGTCAGGCGCTTCATGTCAGCATGGCTGTGAATGCTGTGACCCCGAGAGTGGTTCAGACAC	420
Oy	472	CysProAspLeuGlnLeuArgLeuLeuPheSerArgArgLysGlyLysGlyValAlaGln	491
Db	421	TGCCCCACACTGCGAGCTCAGACGTGCTCTCTCCCGAGGAAAGGAGGTCAAGCTCCAG	480
Oy	492	ValProSerPheArgProGlyLeuLeuThrLeuPheThrHisGlnSerSerTrpProThr	511
Db	481	GTGCCCTCTGTCGGTCCCTACCTCTCTGACCCCTTTTACGACATCAGTCCACGCTGACCA	540
Oy	512	LeuHisGlnCysIleArgValLeuLeuGlyLysSerArgGluGlnArgPheAspProSer	531
Db	541	CTGCACACAGTCATCCGAGTCTGCTGGGCAAGACCGCGGAAACAAGTTTCGACCCCTCT	600
Oy	532	AlaSerLeuAspPheLeuTrpAlaCysIleHisValProArgIleTrpGlnGlyArgAsp	551
Db	601	GCTCTCTGGACTCTCTCTGTGGCTCATCTCATGTTCTCTCGCATCTGGAGGGGGGAC	660
Oy	552	GlnArgThrProGlnLysArgArgGluGlnLeuValLeuArgValGlnGlyProGluLeu	571
Db	661	CAGGCAACCCCGCAGAGCGCGGAGAGAGCTGTCTCGGGGTCCAGAGCCCGGAGTCTC	720
Oy	572	IleSerLeuValGluLeuIleLeuAlaGluIleThrArgSerGlnAspGlyAspIleThr	591
Db	721	ATTCAGCCTGGTGGAGCTGATCTGTGCCGAGGCGGAAACCGGAGCCAGACGGGACACA	780
Oy	592	AlaAlaCysSerLeuIleGlnIleArgLeuProLeuLeuLeuSerCysCysGlyGlyAsp	611
Db	781	GCCGCTGCAGCCTCATCCAGGCGCGGAGCTGCTGCTGCTCAGCTGGTGGTGTGGGAC	840
Oy	612	AspGluSerValArgLysValIleThrGlnHisLeuSerGlyCysIleGlnGlnTrpLysAsp	631
Db	841	GATAGAGGTGCAGAGAGGGGACGGAGCACTGTCTAGGTGATCCACACAGTGGGAGAC	900
Oy	632	SerValLeuGluArgArgCysArgAspLeuLeuGlnLeuTyrLeuGlnArgProGlu	651
Db	901	AGCGTCTGGGCAAGCGCTCCCGAGACTTCTCTCTGACTTACTACCTACAGGCGCGAG	960
Oy	652	LeuArgValProValProGluValLeuLeuHisSerGluGlyAlaAlaSerSerVal	671
Db	961	CTGGGGGGGCGCCGTGAGTCTATGTCACAGCGAAGGGGCTGCGACAGCTC	1020
Oy	672	CysLysLeuAspGluLeuIleHisArgPheIleThrLeuLeuAlaAspThrSerAspSer	691
Db	1021	TGCAGCTGGACCGACACTCATCCACCGCTTATCAGCTCTTTCGGACACCAAGCACTTCC	1080
Oy	692	ArgAlaLeuGlnAsnArgGlyAlaAspAlaSerMetAlaCysArgLysLeuAlaValAla	711
Db	1081	CGGGCGTTGGAGAACCCGAGGGCGGATGCCACGATGGCTGCCCGGAAGCTGGCGGTGGC	1140
Oy	712	HisProLeuLeuLeuLeuArgHisLeuProMetIleAlaAlaLeuLeuHisGlyArgThr	731
Db	1141	CACCGCGTGGCTGCTCAGGACACTGCCCCATATGCGGGCGCTCTGACGCGCGACAC	1200
Oy	732	HisLeuAsnPheGlnGluPheArgGlnGlnAsnHisLeuSerCysPheLeuHisValLeu	751
Db	1201	CACGCAAACTTCCAGAGGATTCGCGGAGCAAAACACACTGAGCTGCTCTCTGACAGTGGT	1260
Oy	752	GlyLeuLeuGluLeuLeuGlnProHisValPheArgSerGluHisGlnGlyAlaLeuTrp	771
Db	1261	GGCTGCTGGAGGCTGTGCAGCGCGCACTGTTCCGACGAGCACACAGGGGCGCTGTGG	1320
Oy	772	AspCysLeuLeuSerPheIleArgLeuLeuLeuAsnTyrArgLysSerSerArgHisLeu	791
Db	1321	GACTGCTCTTGTGCTTCTCATCCGCTGTCTCTAATTACAGGAAGTCTCCCGGCATGTG	1380
Oy	792	AlaAlaPheIleAsnLysPheValGlnPheIleHisLysTyrIleThrTyrAsnAlaPro	811
Db	1381	GCTGCTTCATCAACAAGATTGTGCAAGTTCAATCATTAAGTAACTTACAAATGCCCA	1440
Oy	812	AlaAlaIleSerPheLeuGlnLysHisAlaAspProLeuHisAspLeuSerPheAspAsn	831
Db	1441	GCAGCAACTCTCTTCTGCAAGAACACAGCGCACCGCTCCAGACCTGTCTTTCGACAAAC	1500

OY	832	SerAspIeuValMetLeuLysSerLeuLeuAlaGlyLeuSerLeuProSerArgAsp	851
Dd	1501	ACTGACCTGGTATCTGAATAATCCCTCTTGAGGGGTCAAGCCTGCCAGCAGGGAGCAC	1560
OY	852	ArgTrAspArgGlyLeuAspGluGluGlyGlnGlyGlnSerSerAlaGlySerLeuPro	871
Dd	1561	AGGACCGAACCAGGGCTGGAGCAAGAAGGGGAGAGAGAGCTACGCCGCCTTTGGCCC	1620
OY	872	LeuValSerValSerLeuPheThrProLeuThrAlaAlaGluMetAlaProTyrMetLys	891
Dd	1621	CTGGTCACGCTCTCCCTGTCTCAACCCCTTGACCCGGGGGGAATAGCCCCCTACATTGAA	1680
OY	892	ArgLeuSerArgGlnGlnThrValGluAspLeuLeuGlnGluValLeuSerAspIleAspGlu	911
Dd	1681	CGGCTTTCCGGGGGCCAACAACGCTGGAGGATCTGCTGGAGGTTCTGAGTAACATAACGAC	1740
OY	912	MetSerArgArgAlaArgProGluIleLeuSerPhePheSerThrAsnLeuGlnArgLeuMet	931
Dd	1741	ATGTCCCGGCGAGAACCCGAGATCCTTGAGCTTCTCTTGACCAACCTCAGCGGCTGATG	1800
OY	932	SerSerAlaGluGluCysCysArgAsnLeuAlaPheSerLeuAlaLeuArgSerMetGln	951
Dd	1801	AACCTGGCCCGAGAGATGTTGCCGCAACTCGCCTTCACAGCTTGCCCTCGCTCCATGAC	1860
OY	952	AsnSerProSerIleAlaAlaAlaPheLeuProThrPheMetTyrCysLeuGlySerGln	971
Dd	1861	AACACCCCGCATGTGACGCCCTTCTTGCCACAGTTTCATGACTGCTGGGCGACGAC	1920
OY	972	AspPheGluValValGlnThrAlaLeuArgAsnLeuProGluTyrAlaLeuLeuGlyGln	991
Dd	1921	GACTTGGAGTGTGTGACAGCGCCCTCGGAACCTGCTGAGTACGCTCTCTTGTCGCAA	1980
OY	992	GluHisAlaAlaValLeuLeuLeuHisArgAlaPheLeuValGlyMetTyrGlyGlnMetasp	1011
Dd	1981	GAGCAGCGCGCTGTGCTGCTCCACCGGGGCTTCTGTGGGCGCATGTACGGCAGATGAC	2040
OY	1012	ProSerAlaGlnIleSerGluAlaLeuArgIleLeuHisMetGluAlaValMet	1029
Dd	2041	CCCAACGGCGCATCTCCGAGGCTCGAGGATCTGCATGTGAGAGCGCGTGATG	2094
RESULT 6			
LOCUS	BC010333	2730 bp	mRNA linear ROD 07-AUG-2002
DEFINITION	Mus musculus, clone IMAGE:3601115, mRNA, partial cds.		
ACCESSION	BC010333		
VERSION	BC010333.1	GI:16307573	
KEYWORDS			
SOURCE			
ORGANISM	house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2730) Straussberg, R. Direct Submission Submitted (05-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT	NIH-MGC Project URL: http://mgc.ncl.nih.gov Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amgdbcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., HuiYk, S.W., Hale, S.M., Yoon, V.S., Kowls, C.R., Lawrence, S., Martin, R.G., Mutzny, D.M., Richards, S., Gibbs, R.A.		





Db 83 AAGCCGTTGGTGTGAGTCCCTGCTGACGAGAGAGAGCCCTGAGCT366G 142  
 QY LysProGlyAlaAspGlySerLeuGlnAlaValArgLeuGlyProSerSerGlyLeu 456  
 Db 143 AAGCCGGGGGCGGAGCGTGGAGCGCTGGGCTGGGCGCTGTCGAC36CCCTC 202  
 QY LeuValAspTrpLeuGlnMetLeuAspProGlyAlaValSerSerCysProAspLeuGln 476  
 Db 203 CTAGGGAGACTGGCTGGAATAGTGTGAGCCCGAGGGTGTGACAGCTGCCCGACCTGGCAG 262  
 QY LeuArgLeuLeuPheSerArgArgGlyGlyGlnAlaGlnValProSerPheArg 496  
 Db 263 CTCAGAGCTGCTCTCTCCCGAGAGAGGCAAGGCTCAGCCAGCTGCTGCTGCTGCT 322  
 QY ProTrpLeuLeuPhePheThrHisGlnSerSerTrpProThrLeuHisGlnCysIle 516  
 Db 323 CCTACCTCTGACCTCTCTCAGCATCATGCTGACCTGACCTGACCTGACCTGACCT 382  
 QY ArgValLeuLeuGlyLysSerArgGlnArgPheAspProSerAlaSerLeuAspPhe 536  
 Db 383 CGAGTCTGCTGGGCAAGAGCGGAGACAGAGGTTGACACCTCTGCTGCTGCTGCTGCT 442  
 QY LeuTrpAlaCysIleHisValProArgIleTrpGlnGlyArgAspGlnArgTrpProGln 556  
 Db 443 CTCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502  
 QY LysArgArgGlnGluLeuValLeuArgValGlnGlyProGlnLeuLeuSerLeuValGlu 576  
 Db 503 AAGCCGGGGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 562  
 QY LeuIleLeuAlaGlnAlaGlnIleThrArgSerGlnAspGlyAspTrpAlaAlaCysSerLeu 596  
 Db 553 CTGATCTGCGGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 622  
 QY IleGlnAlaArgLeuProLeuLeuLeuSerCysCysGlyAspAspGlnSerValArg 616  
 Db 623 ATCCAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 682  
 QY LysValIleThrGlnHisLeuSerGlyCysIleGlnGlnIleTrpGlyAspSerValLeuGlyArg 636  
 Db 663 AAGGTGAGCGAGCACCCTGTCAGAGCTGTCATCCAGAGTGGGGAGACAGCTGCTGCTGCTGCT 742  
 QY ArgCysArgAspLeuLeuLeuGlnLeuArgIleGlnArgProGlnLeuArgValProVal 656  
 Db 743 CGCTCCGAGACCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802  
 QY ProGlnValLeuLeuHisSerGlnGlyAlaAlaSerSerValCysLysLeuAspGly 676  
 Db 803 CCTGAGGCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 862  
 QY LeuIleHisArgPheIleThrLeuLeuAlaAspTrpSerAspSerArgAlaLeuGlnIleAsn 696  
 Db 863 CTCATCCACCGCTTCTATCCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 922  
 QY ArgGlyAlaAspAlaSerMetAlaCysArgGlyLysLeuAlaValAlaHisProLeuLeuLeu 716  
 Db 923 CGAGGGGCGGAGTGCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 982  
 QY LeuArgHisLeuPheProMetIleAlaAlaLeuLeuHisGlyArgTrpHisLeuAsnProGln 736  
 Db 983 CTCAGGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1042  
 QY GluPheArgGlnGlnAlaAsnHisLeuSerCysPheLeuHisValLeuGlnLeuLeuGlnLeu 756  
 Db 1043 GAGTTCGCGGACAGACACCTGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1102  
 QY LeuGlnProHisValPheArgSerGlnHisGlnGlyAlaLeuTrpAspCysLeuLeuSer 776  
 Db 1103 CTGCAGCGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1162  
 QY PheIleArgLeuLeuLeuAsnTrpArgLysSerSerArgHisLeuAlaAlaPheIleAsn 796  
 Db 1163 TTCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1222

QY 797 LysPheValGlnPheIleHisLysTrpIleThrTrpAsnAlaProIleAlaIleSerPhe 816  
 Db 1223 AAGTTTGTACAGTTCATCCATAGTATACCTTACCAATAGCCCGACACCACTTCTCTTC 1282  
 QY LeuGlnLysHisAlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuValMet 836  
 Db 1283 CTGCAGAAACGACCGGACCGCTCCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1342  
 QY LeuLysSerLeuLeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArgGly 856  
 Db 1343 CTGAATATCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1402  
 QY LeuAspGlnGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 876  
 Db 1403 CTGGAGAGAGAGGAG 1462  
 QY LeuPheThrProLeuThrAlaAlaGlnMetAlaProTrpMetLysArgLeuSerArgGly 896  
 Db 1463 CTGTTACCCCTCTGACCGCGGCGGAGATGGCCCTTACATGAAAGCGCTTCCCGGGGC 1522  
 QY GlnThrValGlnAspLeuLeuGlnValIleSerAspIleAspGlnMetSerArgArgArg 916  
 Db 1523 CAACCGTGGAGAGATCTGCTGAGGTTCTGAGTACATGACAGAGATGTCCCGCGGAGAG 1582  
 QY ProGlnIleLeuSerPhePheSerThrAsnGlnArgLeuMetSerSerAlaGlnGlu 936  
 Db 1583 CCGAGATCTCTGAGCTTCTTCTGACCAACCTGACAGCGCTGATGAGCTGCGCGGAGAG 1642  
 QY CysCysArgAsnLeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSerIle 956  
 Db 1643 TGTGGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1702  
 QY AlaAlaAlaPheLeuProThrPheMetTrpCysLeuGlySerGlnAspPheGlnValAl 976  
 Db 1703 GAGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1762  
 QY GlnThrAlaLeuArgAsnLeuProGlnTrpAlaLeuLeuCysGlnGlnLysHisAlaAlaVal 996  
 Db 1763 CAGAGCGCGCTCCGAGACCTGCTGAGTACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1822  
 QY LeuLeuHisArgAlaPheLeuValGlyMetTrpGlyGlnMetAspProSerAlaGlnIle 1016  
 Db 1823 CTGCTCCACCGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1882  
 QY 1017 SerGlnAlaLeuArgIleLeuHisMetGlnAlaValMet 1029  
 Db 1883 TCCGAGGCGCTGAGATCTGATATGAGAGCGCGTATG 1921  
 RESULT 8  
 AC102953/c 148996 bp DNA linear PRI 01-JUN-2002  
 LOCUS AC102953  
 DEFINITION Homo sapiens BAC clone RP11-1246C19 from 7, complete sequence.  
 ACCESSION AC102953  
 VERSION AC102953.5 GI:21307577  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 148996)  
 AUTHORS Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 148996)  
 REFERENCE 3 (bases 1 to 148996)  
 AUTHORS Du, H. and Kozlowski, A.  
 TITLE The sequence of Homo sapiens BAC clone RP11-1246C19  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 148996)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission

JOURNAL Submitted (23-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 148996)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 148996)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 6 (bases 1 to 148996)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 7 (bases 1 to 148996)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-JUN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Jun 1, 2002 this sequence version replaced g1:19774560.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sepienewatson.wustl.edu](mailto:sepienewatson.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH1246C19

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7/>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E., Tateo,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is RP11-369D24, 2000 bp overlap; the clone sequenced to the right is RP11-16P10, 2000 bp overlap. Actual start of this clone is at base position 174583 of RP11-369D24; actual end is at base position 63205 of RP11-16P10.

Sequence derived from one plasmid subclone, base position 32899 to 33087.  
 Unresolved tandem repeat from base position 99549 to 100076.  
 FEATURES  
 Source  
 1..148996  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7"  
 /clone="RP11-1246C19"  
 /clone\_1lb="RPCI-11"  
 585..893  
 /rpt\_family="Alu"  
 repeat\_region  
 2518..2618  
 /rpt\_family="MIR"  
 repeat\_region  
 2708..2769  
 /rpt\_family="MIR"  
 repeat\_region  
 2873..3034  
 /rpt\_family="MIR"  
 repeat\_region  
 3347..3438  
 /rpt\_family="L1"  
 repeat\_region  
 3439..3574  
 /rpt\_family="Alu"  
 repeat\_region  
 3575..3857  
 /rpt\_family="Alu"  
 repeat\_region  
 3858..4034  
 /rpt\_family="Alu"  
 repeat\_region  
 4035..4628  
 /rpt\_family="L1"  
 repeat\_region  
 4629..4714  
 /rpt\_family="Alu"  
 repeat\_region  
 4796..5037  
 /rpt\_family="Alu"  
 repeat\_region  
 5039..5296  
 /rpt\_family="Alu"  
 repeat\_region  
 6009..6933  
 /rpt\_family="L1"  
 repeat\_region  
 6938..7061  
 /rpt\_family="Alu"  
 repeat\_region  
 7145..7243  
 /rpt\_family="L1"  
 repeat\_region  
 7256..7559  
 /rpt\_family="MER1\_type"  
 repeat\_region  
 9418..9677  
 /rpt\_family="Alu"  
 repeat\_region  
 9758..9876  
 /rpt\_family="L2"  
 repeat\_region  
 9886..9926  
 /rpt\_family="MER1\_type"  
 repeat\_region  
 9927..10207  
 /rpt\_family="TTCC)n"  
 repeat\_region  
 10208..10220  
 /rpt\_family="Alu"  
 repeat\_region  
 10227..10366  
 /rpt\_family="MER1\_type"  
 repeat\_region  
 11484..11545  
 /rpt\_family="L2"  
 repeat\_region  
 12151..12259  
 /rpt\_family="MIR"  
 repeat\_region  
 12792..13111  
 /rpt\_family="MIR"  
 repeat\_region  
 14332..14451  
 /rpt\_family="Alu"  
 repeat\_region  
 14794..14822  
 /rpt\_family="MIR"  
 repeat\_region  
 14828..15133  
 /rpt\_family="AT-rich"  
 repeat\_region  
 15194..15288  
 /rpt\_family="Alu"  
 repeat\_region  
 16148..16439  
 /rpt\_family="MIR"  
 repeat\_region  
 /rpt\_family="Alu"

```
repeat_region 16468..16637 /rpt_family="Alu"
repeat_region 16647..16944 /rpt_family="Alu"
repeat_region 16972..17244 /rpt_family="Alu"
repeat_region 17254..17442 /rpt_family="L1"
repeat_region 19130..19311 /rpt_family="L2"
repeat_region 19432..19683 /rpt_family="MIR"
repeat_region 19833..19938 /rpt_family="L2"
repeat_region 20105..20243 /rpt_family="MIR"
repeat_region 20951..21003 /rpt_family="L1"
repeat_region 21298..21415 /rpt_family="L1"
repeat_region 21589..21644 /rpt_family="MIR"
repeat_region 21944..22036 /rpt_family="MIR"
repeat_region 22180..22292 /rpt_family="MIR"
repeat_region 22268..23067 /note="CG island (%GC=72.4, O/e=0.80, #CpGs=75)"
misc_feature 22660..22845 /rpt_family="n"
repeat_region 23254..23354 /rpt_family="MIR"
```

Alignment Scores:	
Pred. NO.:	1 41e-186
Score:	3199.50
Percent Similarity:	24.85%
Best Local Similarity:	24.83%
Query Match:	61.05%
DB:	9
	Gaps: 22
	Matches: 148996
	Mismatches: 1010
	Mismatches: 1
	Indels: 6
	Gaps: 3054
	Gaps: 22

US-09-929-769-7 (1-1029) x AC102953 (1-148996)

QY	13	LeuhtirleuAltyPro-----ProArgAlaAsp	21
Dp	136178	TTGGGCGTGGGCCCCGGGTGGAGCTGAAGCGTCACCCGTTGCCTTGCAAGCCGAC	136119
QY	22	AspergilluphegnialaleuleuAspLietrPheproctuguluySproleuProThr	41
Dp	136118	GACAGCGATGCCAGGGCGTCTGGACATCTGGTTTTCCGAGGAGAACCACACCCACC	136059
QY	42	AlapheleuValasptlrserGluAlalaleuleuProaspTrpleuylsleuArg	61
Dp	136058	GCCCTCCCTGGGGGACATCGAAGAGAGCGCTGCTCCTCATCGGCGAAAGCTGGC	135999
QY	62	MettlearSerGluValleuArgleuValAspala-----	73
Dp	135998	ATGATCCGTTCTGAGGTGCTCCGCTGTGTGACCGCCGGTACAGGCTGGGCTTTGGG	135939
QY	73	-----	73
Dp	135938	CAGGGGCTGTACTCAAGGGSTCCCCGGCCCCTTGCTGCTCCTTGtGGGCTtgGGCAGC	135879
QY	73	-----	73
Dp	135878	CTCCGGCTGTGGCAGAGAGAGAAGACTGGCCGGGTAGCCTCAAGGGGATCTGCAAGACATA	135819
QY	73	-----	73
Dp	135818	TGTCTGGCCACAGCAGGCTGTGCTGAGCCCTGGGAGAGAACCTAGGGCACGCTGTGCTG	135759
QY	73	-----	73

Db 135758	GCCCGCTCCGAGAGGGTCTTACTGTCAAGGCTCAGGGGTCTAAGTGTAGAAACAGGTCCTC	135692
QY	73	73
Db 135698	GATGAAAAAAGAGAAAGTGTCTGGGGTTTGAGAGGAGGAGAGAAGTGTGGGCTTC	135639
QY	73	73
Db 135638	TGGGTGGTCCCTGAAAGATTGGGGGTTGTGAGCTGCAGCCGGGGACCCAGCAAGACT	135579
QY	73	73
Db 135578	GAGCAAGGGGCTTGGGCCCAAGTACGTAGTACGTAGTAGGGCTGTGCTGCTCACAGTGAAGAGA	135519
QY	73	73
Db 135518	CACGTGGGGCAGACTTGTGATGATGATGGGGCTGTGCTGTGAAACCATGACTGCCAAGG	135459
QY	73	73
Db 135458	GTTAGGCTGTTGTGTGTGCCCGGAAGCCACCATGTGCTTCGGGGCAAGAGCCCTGAG	135399
QY	73	73
Db 135398	CACCTAGAGAAATAGTCACACAGCAGCTTGGTGCAGCTTGGAAGTACTGGCCCGTGC	135339
QY	73	73
Db 135338	CACCATAGGCCCTTGAGAGCGCCACCCACCCCTGGGGATGGGGCCACACAGCTGCAT	135279
QY	73	73
Db 135278	CCAGGGTGCAAGTGTGTCCGACACGCTGCCCTGTGTGGGCCCAAGTCTCGGGTGG	135219
QY	73	73
Db 135218	GTACCCGAGCCACGTGGGCCAGGGTCTCAGGTGGGGCTGCCGAGTGTGCACCTGCCACG	135159
QY	73	73
Db 135158	GTCTGGGTGGGGGTGCGGAGTGTCACTGTCCCGCAGGCTTGGGTGGGTACACGAGCC	135099
QY	73	73
Db 135098	ACGTGGCCAGGGTCTCAGGTGGGGCTGCCGAGTGTGCACGCCCAGGCTTGGGTGG	135039
QY	73	73
Db 135038	GTACCCGAGCCACGTGGGCCAGGGTCTCAGGTGGGTGTACTGCACACAGCAGCC	134979
QY	73	73
Db 134978	CCAGGTCTCGGGTGGGGTCCGAGTGGCCATTGGCCGGGTTTGGTGGGCTTAAGGAGC	134919
QY	74	89
Db 134918	GGGTGACCCCTGACGCTTCGAGGACTGTGACCCGAGAGGCTGTCTGTTCGTGCATC	134859
QY	89	109
Db 134858	GTTTGGCATCCCGTGTCCAGCATGAGCAAACTCTTCAGTTCTGTGACCAAGGACGTGGC	134799
QY	109	122
Db 134798	CAACGACCCCGAGACTCTGGAGCAAGACATCATGTGACAAAGATTAAGTGCAGGGACCTG	134739
QY	122	122
Db 134738	CGAGTGGGTTCACACACGCTGACCGAGCCCGACCGCCAGGCTTTGGGCGCTGGT	134679
QY	122	122
Db 134678	GCGAGCGGGTGGGTGATCCGAGGTCGGGTGGGTGTGGGTGGCACTCGAGGGCAGGCTTAG	134619



QY	122	-----	122
Db	134618	ACSTGTGGTGGTGGAGGCTGGACAGTCAAGAGGAGCCGAGGCGTGGCCCA	134559
QY	122	-----	122
Db	134558	CCCTGGGGCCAGGACCCCTAGGGCCGTGTTTGAGCTTGGAGGAGCTGTGGGCT	134499
QY	122	-----	122
Db	134498	CAGTGTGACTGAGACTCACCAGCTTGAAGTCCTTGGGAAGTGTGTGACGTTGGG	134439
QY	122	-----	122
Db	134438	GCATCTCCAACTGTGTTTCTGTGGAGGAGGAGGAGCTCTCCCTGGCCCGG	134379
QY	123	-----AsnTyrMetAlaHisLeuValGlnHis	133
Db	134378	CCCAAGTGTGACCTCGTACGCGTTGGCAGATTACATGAGCCACCTGGTGGAGGTCCAGCA	134319
QY	133	sgLuarGlyAlaSerGlyGlnThrPheHisSerLeuThrAlaSerLeuPro	153
Db	134318	TGACGCGGCGCCCTCCGGAGGCCAGACTTTCACCTCTGCTCAGACCTCCCTCCGCC	134259
QY	153	oatArg-----	155
Db	134258	CCGCGAGGACTCTGGGGAGGGGCTGGGCCACAGACAGACCCAGCTGGGAGGGCTGGG	134199
QY	155	-----	155
Db	134198	GCCAAACAGAGACCAGACCTGGGGAGGGGAGCGGGCCCTGCACAGCCACACAGGGAGT	134139
QY	155	-----	155
Db	134138	GCACACTTGTGATGCCACCTGAGAGGGGCCCCACAGAACTCTCAGCTGCTTGGG	134079
QY	155	-----	155
Db	134078	CAGATTTCCTCTGGCCAGCATCAGTTCCCTGTAGAGTCTTGAGATCCTGTGGAC	134019
QY	155	-----	155
Db	134018	CTGAGCAGACTGGGGTCTTGGGGAACAAGAGCCAGGCTGAGCTTCCTCTCTGG	133959
QY	155	-----	155
Db	133958	GTGACAGTCCCCGGGCGAGAGCTGCCATGGCTGTACAGTTCACCCGCTGGAATSCAGGT	133899
QY	156	-----AspSerThrGluAlaProly	162
Db	133898	CAGAAAGTGCACAGCCACTGACATTCCTCCGTCTCGACAGACGACAGAGGACACCCAA	133839
QY	162	spProlySerSerProGluGlnProIleGlyGlnGlyArgIleArgValJalYhrGlnLe	182
Db	133838	ACCAAAAGGACCCAGGAGCGCCCATAGGCGCAGGGCGGATTCGGGTGGGGACCCAGCT	133779
QY	182	uArgValLeuGlyProGluAspAspLeuAlaGlyMetPheLeu-----	196
Db	133778	CCGGGTGCTGGGCCCTGAGGACGACCTGGCTGGCATGTTCTT--CCAGGTAGCGTTGCTGC	133720
QY	196	-----	196
Db	133719	TGCATGTGGGTCTCCACCTGACGCCCTCTCTCGCTGACGCTTCGCTGCCGCTCGC	133660
QY	196	-----	196
Db	133659	ACGACAGCGGCGTTTACTTGTGAGCCGCTGAACCCAGGACAGCTGCACAGCACAGCG	133600
QY	196	-----	196
Db	133559	GCCTTTACACCGCTGACCGCTGACCTGGGACGCTCTTTACAGTTAACCCCCCATGCT	133540

OY	196	-----		196
Db	133539	CTCAGATGGAGGCTAAGCCAAACATTGGGGGAAGGGGGTCGCCCTGCATCTCCTGCCATC	133480	
OY	196	-----		196
Db	133479	CTTGACACACCCTCTGGCGGGGACACC GGTTGTATGTGTAAGAAGCTGAGTTCAGG	133420	
OY	196	-----		196
Db	133419	TCTGCCTGAGGCTGTGACTTTGGAGCGGGGATCTGCAGGCCCTCACACTCTCATG	133360	
OY	196	-----		196
Db	133359	GAGAGCTGGTGGTCCTCGGGGACAGGCTGTGACACACTTTGTTCTGTGGGACACTCCTGG	133300	
OY	196	-----		196
Db	133299	GTTGTTGGGGTCCCTCACTGCTTCATCATACAGGCTCTCAGAGGCCCTGGCCCTCAGGCC	133240	
OY	196	-----		196
Db	133229	ACAAGGTGGGGGCTAACGATGACAGGAGCACAGTGGAGCGGTTTGGACAGGAGCTG	133180	
OY	196	-----		196
Db	133179	CACGGGCGGCTTAGCCTCCCGCTGCGCTCCTCCTCCTCCCTCCCTCCTCCTGACA	133120	
OY	197	-----	-GlnIlePheProLeuSerProAspProArgTrpInsense	211
Db	133119	CTGGCCTTCCCCCGTTCCAGATTTTCCGCGCTCAGCGCGGACCCTGGGGGAGAGCTCCAG	133060	
OY	211	rProArProValAlaLeuAlaLeuGlnGlnAlaLeuGlyGlnIleuValaarValVa	231	
Db	133059	TCGCCGCCCGCTGGCGCCCTGCGCTGCAGAGGCCCTGGGCGAGAGCTGGCCGCGCTG	133000	
OY	231	IcIcIcIySerProGIuValProGIyleThrValArgValLeuGlnAlaLeuAlaThrIe	251	
Db	132999	CCAGGCGAGCCCGCAGAGGTGGCGGCGCATCAGCTCGTGGTCTCGAGGCCCTCGCACCT	132940	
OY	251	uLeuSerSerProHIsIglValAlaLeuValMetSerMetHisArgSerHisPheLeuAl	271	
Db	132939	GCTCAGCTCCCACACAGCGCGGTGCCCTGAGGATGATCATGCACACGATGACCATCTCGGC	132880	
OY	271	aCySProLeuLeuArgIleuLeuCySgInTy-	281	
Db	132879	CTGCCCGCTGCTGGCCACAGCTCTCCACAGTACAGACTTTGGGCGTGGGCTCTTGGCGG	132820	
OY	281	-----		281
Db	132819	GGAGAGGGGACGTAGAGGCTCTGCCCTCCCGAGAGACGGCCGGGTGTCCTGTACTGCTC	132760	
OY	281	-----		281
Db	132759	CTCTGTGGGCTTGTGGGTCTCAGTTTCCAAGTGAGAAAAGACAGTGCCTCCAGGGAGTGGC	132700	
OY	281	-----		281
Db	132699	GCAGGTCCCTTTTTTCCAGTGAACATTCCCGCGGAGCGAGGAGGTGGGGGCAGTGGCGCT	132640	
OY	281	-----		281
Db	132639	CCGCGTGTTCACACACCACAGGCTCGGCGCTCTGCTGCTTCCCATCTGTGCTGTAC	132580	
OY	281	-----		281
Db	132579	CTAGAGCTGTGTTTACTTGTGTGTGTGTGTTTACTTTTCCTTTGAACCGCGTGGAGCC	132520	
OY	281	-----		281
Db	132519	CTCACACAGCTGGGGGTTTGAAGGGGTCTGTGAGAGTGTGTGGGCGCGTGCAGGGGA	132460	
OY	281	-----		281



Db 132459	GGCACCGCTTTTACAGACAGAGAGCGGGGATGAGTTGTGGCCAGAGCTCCAAAGTCCCGGTGGCG	132400
QY 281	-----	281
Db 132399	CAGGCTTCAGCAGCCTCACAGGAGCCCTGCTCTGGGCTCGGGGTGGCGGTCCCGGTGTC	132340
QY 282	-----GlnArgCysValProGlnAspThrG 290	
Db 132339	CTTGCCAGGGTACCCACAGTGCACCTTTGTCGCCGAGCGCTGTGTGCCACAGAGACCG	132280
QY 290	LYPheSerSerLeuPheLeuIlysValLeuLeuGlnMeLeuGlnINTPLeuAspSerProG 310	
Db 132279	GCTTCTCTCTGCTCTTCTCTGTAAGGTGGTCTCTCGAGATGCTGCAAGTGGCTTGACAGCCCTG	132220
QY 310	LYValGlnGlyGlyProLeuArgAlaGlnLeuArgMeLeuAlaSerGlnAlaSerAlaG 330	
Db 132219	GGGTGAGAGGGGGGAGCCCTGTGGGGCACAGCTCAGAGATGTTGGCAGCAGGCTCAGCGG	132160
QY 330	LYArgArgLeuSerAsp-----	335
Db 132159	GGCCGAGGCTCAGTGAATGTTAGAGACCGCGCCGGGGCTGGCAGAGAGCTGGGGGGCACTGG	132100
QY 335	-----	335
Db 132099	ACAGGCGCTCAGTACAGAGTGGCCAAAGCTGCCTGGACACCCGAGGGCTTGGGAGCTGGCTTG	132040
QY 335	-----	335
Db 132039	GGGGGTGGGTACCCGCGTGGGCCAGGGAGACCCCTTCTACCGTGGGGCTCCTCGATGG	131980
QY 336	-----Vala 337	
Db 131979	GATACACCGATGGGGGTGAGGGCTTGAGCCACAGGGCACTCACACAGACCTCCACAGATGG	131920
QY 337	TGGlyGlyLeuLeuArgLeuAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValaValS 357	
Db 131919	GAGGGGGGCTCTCTGCGCTGGCGCAGAGCCCTGGCCCTTCGCTCAGAGACTGGAGTGGTCA	131860
QY 357	erSerThrValArgAlaValAlaIleAlaIlePheLeuArgSerGlyGlnGlnCysSerValaGlnP 377	
Db 131859	GCTCACCGCTCCGTCGCGCTCATCGCACCCCTGAGGTGGGGAGACAGTGCAGCGTGGAGC	131800
QY 377	roAspLeuIleSerLys-----	382
Db 131799	CGGACCTGATCAGCAAAAGTACCTTCTCGGACCATCTCTTGGGCCACAGCGGCGCCTGG	131740
QY 382	-----	382
Db 131739	GAAGACCCCAAACTGAGGCTCTTCTAGCTGCCACATGGGGGAGACGACAGTGTCTCTCA	131680
QY 382	-----	382
Db 131679	GCATGTCCTCCGGAAAGTGCACAGAGTCCCTCCAGATGGCTCAGGGCGGCGCCTTC	131620
QY 382	-----	382
Db 131619	TTGAGAAATGAGTTGGAGCTTCTCTTGGGGGGGCCCTGATCACTGTGCACCCAGCAAA	131560
QY 382	-----	382
Db 131559	GTGAGCCCGCGGCCCGGCGGCACTCTCCAGCCCGGCGGCACTCTCTCCGGGCGCGG	131500
QY 382	-----	382
Db 131499	CTGCATCTCTCTCTGTGGCTCGGGCTGGACTCTGCTTGTGTGCCACACATGCTACCC	131440
QY 382	-----	382
Db 131439	CACACCCACCTGGCTAGGCCACACCCACATCAAGCGGCGCCCAACCGCACTGGGAGTGAC	131380
QY 383	-----ValLeuGlnGlyLeuIleGlnValaArgSerProH 394	

[illegible]

Oy	517	gValLeuLeuEngLySerSerArgIngin	-----	526
Db	130239	AGTCTGCTGGGCAAGAGACCGGAGACAGAGTGGGCGCTTCCTGTCCTCCACACTTGGGCT	-----	130180
Oy	526	-----	-----	526
Db	130179	GCGGCATACTGCGCTTCACACAGAAATCTCCAGACGCGGCGTCTGACTGTATCTCCAT	-----	130120
Oy	527	-----	-----ArgpheAspProSerA	532
Db	130119	GCTGAGACGAGACACCTGCTGCTGTGGAACTTTCCTCCCTCCCAAGATTCGACCCCTCTG	-----	130060
Oy	532	LaSerLeuAspPheLeuTriPalAcysIleHisValProArgIleTyrPglInglYAspG	-----	552
Db	130059	CCTCTCTGGACTTCTCTCTGGGCGCTGCATCATGTCTCTCCATCTTGGACAGGCGGGGACC	-----	130000
Oy	552	LnArgThrPro	-----	555
Db	129999	AGCGCACCCCGCAGCAGAGTCTTGGAGGTTGGGCCATCCGGGAGAGAGGGGGTGCCCATC	-----	129940
Oy	555	-----	-----	555
Db	129939	CGGAGAGGGGGTGCCCATCCGGGAGAGCGCGTGCTGCCGGGCTGATGCAGCGACC	-----	129880
Oy	556	-----GlnLysArgArgGlnGluLeuValLeuArgValGlnGlyProGluLeuIleSerL	-----	574
Db	129879	CGTTACAGAAAGCGCGGAGAGAGTGGTGTCTCGCGGCTCCAGGGCCCGAGACTCATAGCC	-----	129820
Oy	574	eulValGluLeuIleLeuAlaGluIaGluThrArgSerGlnAspGlyAspThrIaIaIaC	-----	594
Db	129819	TGGTGAGACTGATCTCTGGCGGAGCGGAGACGCGGAGACGAGAGCGGGGACACAGCGGCT	-----	129760
Oy	594	YssSerLeuIleGlnAlaArgLeuProLeuLeuSerCysCysGlyAspAspGluS	-----	614
Db	129759	GCAGCTCATCAGAGCCCGGCTGCGCTCTGCTCTGCTCAGCTCTGCTGAGGGAGCATAGAA	-----	129700
Oy	614	erValArgLysValThrGlnHisLeuSerGlyCysIleGlnGlnTyrPglYasp	-----	631
Db	129699	GTCGTACGAGAGGTGACGAGAGACTGTTCAGGCTGCAATCCAGCATGGGGAGACAGTGTCT	-----	129640
Oy	631	-----	-----	631
Db	129639	CAGGCCCCCGGCCCCAGATGCTGCTGGGAGTGAAGCCATGAGAGGGGTCGAGGCCACG	-----	129580
Oy	631	-----	-----	631
Db	129579	CCTGGGTGGGTTCCCATTCAGCGCTGTGTCCTCCACCCACCCCTGTACTGGGGCCCCGT	-----	129520
Oy	632	-----SerValLeuGlnYArgArgCysArgAspLeu	-----	641
Db	129519	CCGCGCTGGTGAAGCCCTGCTCACCTCTTCACAGAGGTGCTGGGAGGGGCTCCGAGACCTT	-----	129460
Oy	642	LeuLeuGlnLeuTyrLeuGlnArgProGluLeuArgValProValProGluValLeuLeu	-----	661
Db	129459	CTCTCGACGCTCTACCTACACACGCGGACGAGACTGGGGGTGCCCTGCTGAGTCTACTG	-----	129400
Oy	662	HisSerGlnGlyAlaIleAspSerSerValCysLysLeu	-----	674
Db	129399	CACAGCGAAGGGGCTGCCAGCAGCAGCGTCTGCAAGST -GAGGGCGCACTCTGCCGCTC	-----	129341
Oy	674	-----	-----	674
Db	129340	CCGCGCGCGCTGCCAGGTGCGCTGGCCGGGAGAGGAGGGGGAGGCGCACAGTGTAC	-----	129281
Oy	674	-----	-----	674
Db	129280	ATGGCCATCGCGCGCTACCCCTTGTGCCCTTCCAGTCCGCGCGCCCTGTGGCCCTGT	-----	129221
Oy	674	-----	-----	674
Db	129220	GGGTTTGTCTCTGTTAGATTTTATACCGAGAGACTTGGCGGTGCTCTGCTGCGACAGTC	-----	129161

OY	674	-----	674
Db	129160	TGGAGAGCTGCGTGTGTGGGCACTGGGTGGCCTGGCATTGTGTGATGCGCCGTTGGT	129100
OY	674	-----	674
Db	129100	TTATCTGTTCACTGCTGTGTGGACGTGGGTGCAGTTGTTGGGCACTGACAGTGC	129040
OY	674	-----	674
Db	129040	TGGGACTGCTGCCCCGGGTCTTGGCGGACGGCTGCTCTCTCTCCAGTGACCTGG	128980
OY	674	-----	674
Db	128980	TGGGAGAGCTGGGCGTCCGCTACCGGCTTCCCGACCGCGGAACTGCAAGCCTGT	128920
OY	674	-----	674
Db	128920	CTTCTGCTGTACCAACAGAGGAGGAGCTGACTTTCCTAATCAATTCTGCTGTGTTTTT	128860
OY	674	-----	674
Db	128860	TTCACTTATATGTCATTTTATTTTTGAATCACACTTGATGTGCAGTTTGTGGAGGCT	128800
OY	674	-----	674
Db	128800	TTTTGAGGCGAGGAGTGTCTGTTGCCAAGGCTGAGTGCAGCGGTACATCATCTCACTG	128740
OY	674	-----	674
Db	128740	CAGCTCAACCCCTGTGCTCAAGTGAAGCTCCACACTCAAGCTCCCGAGTACGTGGAT	128680
OY	674	-----	674
Db	128680	CACAGGACAGTGCACACACACACACACTTACTTTTGTATTTTGAATAAGAGAGGCT	128620
OY	674	-----	674
Db	128620	ATTTTCAATTGCCCAAGCTAGTCTTGAACTCCTGGGCTCTGTGATCTTCTGCCGCGCT	128560
OY	674	-----	674
Db	128560	CACAGAGTGTGGAGTGGACAGGCGTGAGCACACTGCGCTGCTGGCCGATGTAGTTT	128500
OY	674	-----	674
Db	128500	GGCTTGTTCGCCCTCATTTGATCAATTGTGTGTGAGCTTGTCTTATCTGTATGGG	128440
OY	674	-----	674
Db	128440	TTTTCCAGAGGCTGGGTGTGGCTGGCTGGCTCCGAGGTGCTCATGAAACCCGTAG	128380
OY	674	-----	674
Db	128380	GAGCAGTTCCTGTCTCATGCGTGCACGCGGTGATCAGGCGCACAGCTGGCGCCA	128320
OY	674	-----	674
Db	128320	GGCTGAGTGTGGGTGGGCAAGTGAAGGGAGCCAGCGGCCCGGCTCTAGAGCCCTCC	128260
OY	675	-----	675
Db	128260	CTGACGTGAGGAGCATCATCCACCGCTTCATCAGCTCTCTGGGACACACGACTCC	128200
OY	682	-----	682
Db	128200	CGGGGTTGGAGAACCGAGGGGCGGATGCCAGCATGGCTGCCGGAAGCTGGCGGTGGC	128140
OY	712	-----	712
Db	128140	CACCGCGCTGTGCTGCT--CAGGTCTGTCCAGAAAGCCCTGCGCCAGACACTGGTT	128080
OY	717	-----	717

Db 128081 TAACAGGGGGTCTGGGCTCTCCAGCCTGGGCAACAAGACAATACTCCAAAAACAAC 128022  
Qy 717 ----- 717  
Db 128021 AAAAAAGCATGTTCCATGGGCTGAGCTTGACCGGGGGAAGTCAAGACTGAGCTGGGC 127962  
Qy 717 ----- 717  
Db 127961 CTCAGGTGGAGAGCTGGTGGGGCTCAGAGAGGGGATCTCATCCCACTCTCTAG 127902  
Qy 717 ----- 717  
Db 127901 GGCAGAGAGGTGTGGGGGGGAGACAGGGGCTTGCTGCCGGGGCAGAGTGGGCT 127842  
Qy 717 ----- 717  
Db 127841 TCTGGGGAGCTGGCCAGGTGAGCCGCGGGGGCTGACTGGGCTTACCGCGCC 127782  
Qy 718 ---ArgHisLeuProMetLeuAlaLeuLeuHisGlyArgThrHisLeuAspPheGln 736  
Db 127781 CGCAGGCACTGGCCCATGATGCGGGCTCTGCAAGGGCCGACCCACTTCCAG 127722  
Qy 737 GluPheArgGlnGlnAsnHisLeuSerCysPheLeuHisValLeuGlyLeuLeuSLeu 756  
Db 127721 GAGTTCGGCAGCAAAACACCTGAGCTGCTTCTGACAGTGTGGGCTGCTGAGCTG 127662  
Qy 757 LeuGlnProHisValPheArgSerGlnHisGlnGlyAlaLeuThrPaspCysLeuLeuSer 776  
Db 127661 CTGAGCGCGACGTGTTCGGAGCAGCAGCGGGGGCTGTGGACTGTGCTTCTGTC 127602  
Qy 777 PheLeuArgLeuLeuLeu----- 782  
Db 127601 TTCAATCCGCTGCTGCTGTGTGTGTCTCCGCGCGCTCGGCTCTCCCTCTACTCC 127542  
Qy 782 ----- 782  
Db 127541 CTGTCTCTGGCCAGGCTGCTCCTACCCACTCGCGAGAGACTGGAGACTCTGCAG 127482  
Qy 782 ----- 782  
Db 127481 CCAAGGCCCAAGGCAAGCCACTGCTGGGGACCAAGTCCACCCCTGGAAAGCCCTGCC 127422  
Qy 782 ----- 782  
Db 127421 CTGCTGCCCTCAGCTGCTGGGGGGCCAGCTCTTGAAGACCCAGACCCAGACCTCC 127362  
Qy 782 ----- 782  
Db 127361 GTGCTGTCTCTGGGTGTGCTGTGTACCTCGAGCCCTCATCATCTGGGACATGC 127302  
Qy 782 ----- 782  
Db 127301 CCCAGACAGCATGTCCACACCTGACAGGCCATGCCGGACACCTTGGGCTGTGACGC 127242  
Qy 782 ----- 782  
Db 127241 CCCAGTGGTGGCAGAGGGGGCTGTGGCTGTGGGGTCCATGCCCTAGGGCTGGGAG 127182  
Qy 782 ----- 782  
Db 127181 CAGATGGAGCAGAGCTCTGTTTTGGCCCTGCGGGGAGAGCACTGAGGGTGGCCCT 127122  
Qy 783 -----AsnTyrArgIysSerSerArgHisLeuAlaAlaPheLeuAsnIysPhe 798  
Db 127121 TTTGCTTTGCAGAAATTAAGGAAGTCTCCCGCCATCTGGGCTTATCAACAAGTTT 127062  
Qy 799 ValGlnPheLeuHisLysTyrIleThrTyrAsnAlaProAlaAlaIleSerPheLeuGln 818  
Db 127061 GTGAGGTTCATCATTAAGTACATTACACATGCCCCAGAGCATCTCTCTCTGAG 127002  
Qy 819 LysHisAlaAspProLeuHis----- 825  
Db 127001 AAGCAGCCGACCCGCTTCCAGTGAAGTCCACAGGGCTCAGAGGAAGCTCGGGCTCCGG 126942  
Qy 825 ----- 825  
Db 126941 CGAGGGTCCAGCGGGGTCTGGAGGGTGTCTCTGGGAAGGCGGTGTCTGATATGCT 126882  
Qy 825 ----- 825  
Db 126881 GCTCTGAGGCCGGGAGGGGCTGTGAGCGTGTGTCTATCCAGAGGGCGAGAGCCGCC 126822  
Qy 825 ----- 825  
Db 126821 CCACCTGTGCCGAAGCCGTGTGAACCCACCCCTGACCTCTGTCTGTCCGCT 126762  
Qy 825 ----- 825  
Db 126761 CCAGCAGATGTGGCTTCCCGGCCGCCCTCTCTGAGAGTCTGGGCTCTCTGAT 126702  
Qy 825 ----- 825  
Db 126701 GCCTGTGCCACAGCTCCCATGCCACCCCTAGTTGACCCCTGACCCAGCCTTCTCTG 126642  
Qy 826 -----AspLeuSerPheAspAsn 831  
Db 126641 CCAGGCTCTCTCCAGGTGAAGACATCTTCTCTCCAGGACCTGCTCTTCCAGAAC 126582  
Qy 832 SerAspLeuValMetLeuLysSerLeuLeuHisGlyLeuSerLeuProSerArgAsp 851  
Db 126581 AGTGACCTGTGTATGTCTAAATCCCTCTGTGAGGGCTGACCTGTCCAGAGGGAGCAG 126522  
Qy 852 ArgThrAspArgGlyLeuAspGlnGlyGlu----- 862  
Db 126521 AAGACCAAGCAGAGGCTTGAGCAAAAGGGCGAGGGTGAAGTGGGGGTGCCAGGGGCTGCTG 126462  
Qy 862 ----- 862  
Db 126461 GAAGCAGCTGGGGAACAGTCTACAGGCGCTGGTTTTCCAGCGACAGCCAGCCCTGGG 126402  
Qy 862 ----- 862  
Db 126401 CCGTCCAGGGGCTCCGATGGGGGCTGTGTGAGCCTGGGACCTTGTGCTGCCCCCTGA 126342  
Qy 862 ----- 862  
Db 126341 CCCACCTCAAGGGGAGCAAGACAGTGAAGTGTCTGACAGAGCATCAGACACATCGCCC 126282  
Qy 862 ----- 862  
Db 126281 CCGGACATCGCCCGCCCGGCAATGACCCTCCCAACCCCGCTGCTGGACATCC 126222  
Qy 862 ----- 862  
Db 126221 ACCATCCACAGTGTGTGTGCGAGGTGCTCTCCAGAGAGCTCCCTGTCTACTG 126162  
Qy 862 ----- 862  
Db 126161 GGCAGAGCTTGTGTTTTTCCCTGCTGAGGTGACAGCGACCGACAGCTCTCT 126102  
Qy 862 ----- 862  
Db 126101 CCTGTGCCCGGAGAGAGCGCCAGGCCGGGTCCCAAGGGGCCAATCCGCTTGTGCCCA 126042  
Qy 862 ----- 862  
Db 126041 GTGGGTCTGGGCCCATTTGGGAGAAACCCAGAGGAGAGACTGGCAGAGTGCATGCG 125982  
Qy 862 ----- 862  
Db 125981 CCGATCCAAAGGACAGTCCCGGCTCTGTGCTCACCCCATTTGTGAGCGCTGTGT 125922  
Qy 862 ----- 862  
Db 125921 GCAGGCACTTGGAGAGCTGGGCAAGTGGCCGCGGTGTCTGCGGAGGGCAGGTCTGTG 125862

QY 863 -----G1 863  
 Db 125861 CCTCAGTGGCAGCTCCGCGCTCCGCTCCGACACTGCCACTGCTCTCCAGA 125802  
 QY 863 ucJuserSerAlaGlySerLeuProLeuValSerValSerLeuPheThrProLeuThrAl 883  
 Db 125801 GGAGAGCTCAGCGCGCTCTGCGCTCCGCTGACGCTCTCCCTGTACACCCCTCTACCGC 125742  
 QY 883 aAlaGluMetAlaProTyrMetLysArgLeuSerArgGlyGlnThrValGlu----- 900  
 Db 125741 GCGCGAGATGGCCCCCTACATGAACGGCTTTCGCGGGCCAAACGGTGAAGGATGAGTC 125682  
 QY 900 ----- 900  
 Db 125681 AGGCCCTGCTTCAACCCAGCCAGGTCAGCGCCCTGGGTCCCATGGGGCTCGGGGACAG 125622  
 QY 900 ----- 900  
 Db 125621 CCAGTAGCCATTGGGCACTGACTTTGGCGTGAAGCCTGGGCCACGCTGGGCTGTGGGTGG 125562  
 QY 900 ----- 900  
 Db 125561 GTGGGCTGTCTCTACAGTGGGAGACACCCCGAGAGAGGCGCTTAGAACTCAG3GGGCC 125502  
 QY 900 ----- 900  
 Db 125501 AGGGAGCTGGAGGTGGCTGCTCTGCTTTCAGGACAGCAAAAGAGGAGGCCCTTAA 125442  
 QY 900 ----- 900  
 Db 125441 GGTGTGGGGGTACAGCAGAGAAAGTCAAGCGGGTGGGCAAGACAGCGGTGGGAGT 125382  
 QY 900 ----- 900  
 Db 125381 ACCGGTGGCCGCTGTCAAGAGAGACCCAGGTCGAGCGCGCGAGTGGAGTCCG3CCGTG 125322  
 QY 900 ----- 900  
 Db 125321 GGGACACTGTGCTAAGACACCTTGATATGGGGCTGCCGTGAGGGGTCAAC3CCCTCT 125262  
 QY 900 ----- 900  
 Db 125261 CCTTGGAGCCACTAGTAGGGGAGGCGCGGATTTGTGAGACTTGGGGCAG3TGACT 125202  
 QY 900 ----- 900  
 Db 125201 GCGTGGGGCTCACTGGGCTCATGCTTTCCAGTGTCTTCTGTGTGCTC3GGGCC 125142  
 QY 900 ----- 900  
 Db 125141 TTGCCCCCTCTGGGTGGCTGGGGTGGGAGAGGTCGCTGGCCCTTACGTTCT 125082  
 QY 901 -----AspleuLeuGluValLeuSerAspIleAspGluMetS 913  
 Db 125081 GGTCTGGCCCTCTCTCTGTCTGATCTGCTGAGAGTTCTGAGACATAGACAGAGATGT 125022  
 QY 913 eArgArgArgProGluIleLeuSerPhePhe----- 923  
 Db 125021 CCGGGGAGAGACCCGAGATCTTCTTCGCTGAGCTGAGTGGGGCTGAGGGAGA 124962  
 QY 923 ----- 923  
 Db 124961 GCTGGAGAGACATCTCCCTTGGGGAACCCAGACAGGGCCACCAACCATGGGTGC 124902  
 QY 923 ----- 923  
 Db 124901 CCGCGCTAGACCCAGAGAGTCTCTGAGGCTTGGGGTGGGGCCCAAGCCAGG 124842  
 QY 923 ----- 923  
 Db 124841 TGTTACAGTCAAGCGCGCCAGAGAGGCTGCTTACGCCCGGGGCTGGAGCGGCC 124782

QY 923 ----- 923  
 Db 124781 AGTGGACGTGCTCTGCTCTGATAGGGGCTTAGGCTCCCTGCGCTTGCCGCCAAC 124722  
 QY 923 ----- 923  
 Db 124721 CATCAGAGACCAAGAGTGAAGGGGCTTGAATGGGGGCTGAGACCCCTTGTCAACACC 124662  
 QY 924 -----Ser-ThrAsnLeuGlnArgLeuMetSerSerAlaGluGlyCys 937  
 Db 124661 TTCCACCTTTCCCGGTGACAGCAACCTGACCGGCTGTATGAGTCCGGCCGAGAGTGT 124602  
 QY 938 CysArgAsnLeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerPro----- 954  
 Db 124601 TGCCGACACTGCGCTTCAAGCTTGCGCTGCGCTCCATGACAGAACGCCAGGTGAGGC 124542  
 QY 954 ----- 954  
 Db 124541 CACCGCGGCTCTGTCCCGCGCCAGCCGCTGGGGCCCTGTAAACCAGTCAACCGCTGA 124482  
 QY 954 ----- 954  
 Db 124481 CCGCGCTGTGGCCAGACAGGCTTCCAGCGGCTCAGTCGCGCTGTGGGGGCTC 124422  
 QY 955 -----SerIleAlaAla 958  
 Db 124421 TGCGTGGGACAGAGGTGAGGTTCCCACTGAAGTGTGGGCTTTTCAGCATTTGCAACC 124362  
 QY 959 AlaPheLeuProThrPheMetTyrCysLeuGlySerGlnAspPheGluValGlnThr 978  
 Db 124361 GCTTCTCCGCCACAGTTCATGATCTGCTGGGACAGGACACTTGTGAGGTGGCAGAGC 124302  
 QY 979 AlaLeuArgAsnLeuProGluTyrAlaLeuLeuCysGln----- 991  
 Db 124301 GCGCTCCGAACTGCTGATGAGCTCTCTGTGCCAAGTGAAGACACTGGCCCCGCC 124242  
 QY 991 ----- 991  
 Db 124241 CAGGGCCCTGGCAAGCCCTGGGGAGGCGGAGGTGGCTCCCGAGAGGCTGGCTTGGG 124182  
 QY 991 ----- 991  
 Db 124181 AGTCCGAGGCTGTGCTACTGCGGCGAGAGTGTGTGGCCAGGCTGACGTGAGGGCCC 124122  
 QY 992 -----GluHisAlaIleValLeuLeuHisArgAlaPheLeuValGlyMetTyr 1007  
 Db 124121 GTTTCGCCACAGAGACAGCGGCTGTGCTGCTCACCGGCTTCTGTGGGCAATGTAC 124062  
 QY 1008 GlyGluMetAspProSerAlaGlnIleSerGluAlaLeuArgIleLeuHisMetGluAla 1027  
 Db 124061 GGCAGATGAGACCCAGCGGCGAGATCTCGAGGCGCTGAGATGTGCATATGAGAGCC 124002  
 QY 1028 ValMet 1029  
 Db 124001 GTGATG 123996  
 RESULT 9  
 AC130221  
 LOCUS 219964 bp DNA linear HTG 08-AUG-2002  
 DEFINITION Mus musculus chromosome UNK clone RP23-168B1, WORKING DRAFT  
 AC130221  
 ACCESSION AC130221  
 VERSION AC130221.1 GI:22138681  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 219964)  
 AUTHORS McPherson,J.D. and Waterson,R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 219964)

AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-AUG-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc/index.shtml>  
 Contact: [submissions@wustl.wustl.edu](mailto:submissions@wustl.wustl.edu)  
 Project Information  
 Center project name: M\_BA0168E11

----- Summary Statistics -----  
 Sequencing vector: M13; 0%  
 Sequencing vector: plasmid; 100%  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 217436 bases at least Q40  
 Consensus quality: 217966 bases at least Q30  
 Consensus quality: 218400 bases at least Q20

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 2186: contig of 2186 bp in length  
 \* 2187 2286: gap of unknown length  
 \* 2287 6318: contig of 4032 bp in length  
 \* 6319 6418: gap of unknown length  
 \* 6419 33825: contig of 27407 bp in length  
 \* 33826 33925: gap of unknown length  
 \* 33926 80048: contig of 46123 bp in length  
 \* 80049 80148: gap of unknown length  
 \* 80149 128323: contig of 48175 bp in length  
 \* 128324 128423: gap of unknown length  
 \* 128424 219964: contig of 91541 bp in length.

FEATURES  
 source  
 1. 219964  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="YUNK"  
 /clone="RP23-168E11"

misc\_feature  
 /note="assembly\_name:Contig13"  
 2287..6318  
 misc\_feature  
 /note="assembly\_name:Contig14"  
 6419..33825  
 misc\_feature  
 /note="assembly\_name:Contig15"  
 33926..80048  
 misc\_feature  
 /note="assembly\_name:Contig16"  
 80149..128323  
 misc\_feature  
 /note="assembly\_name:Contig17"  
 128424..219964  
 /note="assembly\_name:Contig18"

BASE COUNT 51706 a 56458 c 58030 g 53268 t 502 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 8,96e-154 Length: 219964  
 Score: 2673.50 Matches: 844  
 Percent Similarity: 27.70% Conservative: 75  
 Best Local Similarity: 25.44% Mismatches: 93  
 Query Match: 51.01% Indels: 2317  
 DB: 2 Gaps: 22

US-09-929-769-7 (1-1029) x AC130221 (1-219964)

OY 18 PROARGALASPSPSERGLUPHEGLNLALEULEUASPILETRPHPHROGLULUYS 37  
 |||||  
 Db 155249 CCTAAGCTGGTGACAGGAGTGTAGTGTGCTGGACATCGTTCCAGAGAAG 155308  
 OY 38 PROLEUPROTHRALAPHELEUVALASPHRSERGLUGLUALALEULEUPROASPTP 57  
 |||||  
 Db 155309 CCTTTGCCACAGCCCTTCCTGTGTGACACAGTCGGAGAGGCCCTGCTGCTACCTGG 155368  
 OY 58 LEULYSLEUARGMETILARSGSERGLUVALLEUATGLEUVALASPALA----- 73  
 |||||  
 Db 155369 CTGAAGCTGCGAATATATGCTGTAGGTCGCCGCTTGAGAGC--TGTTACAGTGAG 155427  
 OY 73 ----- 73  
 Db 155428 GGTAGGGAGGGGCTCGGCGCTACCTGCTGCCGGGGAGCTTAACCTGAGGTGCACA 155487  
 OY 73 ----- 73  
 Db 155488 CCTTCAGGACTCTGAGCGCTCAGAGACAGGCTGAGAGAAAAGGCTATNAGAGACC 155547  
 OY 73 ----- 73  
 Db 155548 TGGCTTTAGAGGACAAAGATAGTTCAGAGGCTTTGGAGAACATGTGGCTATAGGC 155607  
 OY 73 ----- 73  
 Db 155608 AAGCTGCAGTACGACAGAGGAAGTCCATCGGGGTAGGCGTGTGCCAGCTTGGGTCTT 155667  
 OY 73 ----- 73  
 Db 155668 CAGTCTGGCCCCAGCAGGAAAGAGACATCCATGCTGTGGTGTGAGCTGGGTAAGCC 155727  
 OY 73 ----- 73  
 Db 155728 AGTGAAGACATACCATGTAGTGTGACTTAGTGGCCAGAGGTCATGTGCATGTCA 155787  
 OY 73 ----- 73  
 Db 155788 TTGCTAAAGAACCATGGGGGACTACATATTGGCAGAGCCAGGCTTGTGCTGCACC 155847  
 OY 74 -----ALAleuGln 76  
 Db 155848 TGAGATACCTTGCAGAGCAGCGCCCGAGCTGAGCATCTTTGCCACACCCCTGCAG 155907  
 OY 77 ASPLLEUGLUPROGLNGLLEULEULEUPHEVALGINSERPHGLYILEPROVALSERSE 96  
 |||||  
 Db 155908 GACCTGGAGCCCCACAGCAGCTGCTCTTCTGTCAGTCTTTGGCATCCCCGTGCCAC 155967  
 OY 97 METSERLYSLEULEUGLNPHELEUASPGINALAVALAHLSASPPIROGLINTRLEUGLU 116  
 |||||  
 Db 155968 ATGACCAAGCTGCTGCAGACTGTGACAGCGGTAGGCCAGGACCCCAAAACCTGGAG 156027  
 OY 117 GINASNIIEMETASPLYS----- 122  
 |||||  
 Db 156028 CAGAACATCATGACAAAGATNAGCCCTGGGGCCGCGAGCTGGCAGACCCCTGAGGCC 156087  
 OY 122 ----- 122  
 Db 156088 CACCTAACCCCTCTAGGCGAGATGTGTGCTCTCGGGTCAGTGGGCACCTACCACTGTG 156147  
 OY 122 ----- 122  
 Db 156148 GCCAAGCTCCAGAGCTGTGAGACTGACTGTGCGCATCACTACTGAGCTGAGAGA 156207  
 OY 122 ----- 122  
 Db 156208 GACCCAGCAGCTCAGTCCCTGCGAGCTTTGTTTGTGACCTTTTGTGACACAGTGT 156267  
 OY 122 ----- 122  
 Db 156268 TCTCTGTAGCCTTGGCTATCTCAAACTAGCTTTGTAGACCAAGCTAGCCTCAACTC 156327  
 OY 122 ----- 122

Dh	156328	AGAGATCTTGCCCTGCTGCTGCTCTTAAGTGTCAACAGACTTAAGGTGTGTGTCCACCACTGCCCA	156387
Qy	122	-----	122
Dh	156388	GCTAGCTCTTCGACGAGCATAGAGAGACTCGGTGGCATGTGGTGTGTCCAGCCGCCCGG	156447
Qy	122	-----	122
Dh	156448	AGTCTCATGTTTGGTGTGAAAAAGCAGGCTGGCTTCACCCACCCCAAGTGTCCCTGGGTG	156507
Qy	123	-----AsnTyrMetAlaHisLeuValGluValGlnHisGluArgLysSerG	139
Dh	156508	CTGTTTTCAGATATTACATGTGGCCCACTGTGTAGGTACAGACAGAGAGAGTGTCTGTG	156567
Qy	139	-----yGlyGlnThrPheHisSerLeuLeuThrAlaSerLeuProProArgArg-----	155
Dh	156568	AGGCCAGACTTTCACACTCCTCCACAGCCTCCCTCCACCCCGAGAGGTACAGCCCC	156627
Qy	155	-----	155
Dh	156628	CTCCCTATCCCTACATCTTACTCCCGAGCTGGACCGGCTTTTGCCCTCATGTGTTC	156687
Qy	155	-----	155
Dh	156688	CTGCTTGTGCGACACTTATTTTGGGGGCTCTTATGTGGCCCAAGTCCAGCCCTGCTT	156747
Qy	155	-----	155
Dh	156748	CTGGTCACAGTCGCTGGCAGGACAGCTGTGGCAGAGCTGTGCGCAAGTTTGAAGCC	156807
Qy	156	-----AspSerThrGluAlaProLysProLysSerSerProLug	169
Dh	156808	ACACTTCTCTGTCTCTGCACACAGCAGCGAGGCTCTTAAGCCAGAAAGACGCCCTGAGC	156867
Qy	169	-----LeuProLysGlnGlyArgLysValGlyThrGlnLeuArgValLeuGlyProGluA	189
Dh	156868	CCCCACACGGCCAGGGAGGAGGACTGGCGAGGAGGACGAGGTGCGGCTGTGGCCCTGAGG	156927
Qy	189	-----SspAlaLeuAlaGlyMetPheLeu-----	196
Dh	156928	ACGACTTACGTGGCATCTTCTCT--ACAGGTACAAAGGCTACCGTTCAGAGGCCCTCAGCA	156986
Qy	196	-----	196
Dh	156987	GCACCGCCTGCACATGTCTGCCCTGCATGTGTACTCGTACATGTGTCTGGCCACTATT	157046
Qy	196	-----	196
Dh	157047	CCATTCTTATCTGTGTGCTCACTCTTCATGTCACTGTATGTAAGTGTGTGGGACCCC	157106
Qy	196	-----	196
Dh	157107	TGTGTGTCTCTGT	157166
Qy	196	-----	196
Dh	157167	TAGACTCCCTGCGCACATCATCCAGCAAAAGATCTGGGGTCCCTTAGTTAATGACACTGGGA	157226
Qy	196	-----	196
Dh	157227	CGGAATCTGATGGGCTGTGAGCGAGGCTACGCGATGGGAGCCACGGGGTCCGGGGCTCTCG	157286
Qy	197	-----GlnIlePheProLeuSerProAspProArgTyrPglInsers	210
Dh	157287	TTCCTTATCTTCTGTCTCTCCAGATCTTCCACACTGACGCCGAGACCCAGGTGTGACAGCT	157346
Qy	210	-----erSerProArgProValAlaLeuAlaLeuGlnGlnAlaLeuGlyGlnGluLeuAlaArgV	230
Dh	157347	CCAGGCCACGCCCTTGTCCCTGTGCACATGCGACCAAGCCCTGTGGCCAGAGAGTGGCTGTG	157406
Qy	230	-----aIValGlnGlySerProGluValProGlyIleThrValArgValLeuGlnAlaLeuAlaT	250

Dh	157407	TCGCCAAGGGAAACCCCGAGGTGGCAGGATCAACAGTCCGCTTTCGACGACATGACA	1574666
Oy	250	hrLeuLeuSerSerProHisGlyValAlaLeuValMetSerMetHisArgSerHisPheL	270
Dh	157467	CCCTGCTGAGCTCCCCCGACGCTGGCACCCCTGCTCTGCTGATGACATCAACAGCACTTCC	157526
Oy	270	euAlaCysProLeuLeuArgGlnLeuGlyGlnIlyrGlnArgCys-----	284
Dh	157527	TTTCTGCGCACTGATGCGCGAGCTTACACAGTACACAGTATGCTGGACAGTGGCTGGC	157586
Oy	284	-----	284
Dh	157587	GGGAGGCAAGGGCAGTGGAGACTAGTGGGCCAAGTGTGCATGGCTCGCTACTGTCCAC	157646
Oy	284	-----	284
Dh	157647	CCTCAGCGCTGTGTGTGTTCACACATGTCTGAAGGCTTCCCTTAGCGGTCGTATGAT	157706
Oy	284	-----	284
Dh	157707	TTCTGTCTTTTCACACAAAAAAGAGGGGAGAGACAGCAAGGGCAAGGCCATATCG	157766
Oy	284	-----	284
Dh	157767	TAGAGGCGATGATGATGATGAGCTTTGTACACCCCTGTGATGATGACAGAGTGTCTGTGT	157826
Oy	284	-----	284
Dh	157827	ACGTGCTGT	157886
Oy	284	-----	284
Dh	157887	TGGGTATACCCATGAAGCACAGCCTCAGAGACTGTGTGTTAAGCCCTTCAGCCTATGCT	157946
Oy	284	-----	284
Dh	157947	GGCATTTAAGGAAATGCTCTCTTGAAGTGGCAAGAGATGATGTTGGATCTCTGGCCCTGGC	158006
Oy	284	-----	284
Dh	158007	CCTGACCTAGGGCCTGTGTCTGTGGCTGAGACTGCCCCCAGGAGGAGGAGGCTGTGGAGTTC	158066
Oy	285	-----ValProGlnA	288
Dh	158067	CTGTGCCAGCAGTGGGCTCAATGTGGCTGTGCCCTCCACAGCGCTGTGCCCGGAGG	158126
Oy	288	sPThrGlyPheSerSerLeuPheLeuIysValLeuLeuGlnMetLeuGlnTrpLeuAspS	308
Dh	158127	ACACTGCGTTCCTCGCTCTCTTCTTAAGTGCCTCATGACAGATCTCTGAGCTGTGATA	158186
Oy	308	erProGlyValGlnGlyGlyProLeuArgAlaGlnLeuArgMetLeuAlaSerGlnAlaS	328
Dh	158187	GCCCTGCTGTGAGAGATGGGCCCTTGCAGGCCCAACTCAAGTATTTGCTACTCGCTATT	158246
Oy	328	erAlaGlyArgArgLeuSerAspAla-----	336
Dh	158247	CAGCAAGGCAAGATCAGCGACGCG--TGAGCACTGAGCATGACACGCTGAGGGTGGCGGC	158305
Oy	336	-----	336
Dh	158306	GGGCGGGAGACTGTGGCGCTTACTGTGTGTGGAGGCAAGCCTTCAGAGAGCGGTGTAGTTC	158365
Oy	336	-----	336
Dh	158366	TTCTGAAGAACAGCAGCAGCTGGCAGAGAGCGTGGCTGTGATGTGTGGCAAAACGAGAG	158425
Oy	337	-----ArgGlyGlyLeuLeuArgL	343
Dh	158426	CTGTGTGTGTGTGCATGTGTTTGTACACTGCTGTGACGTGCGAGTGTGGGCTCTGTGACT	158485
Oy	343	euAlaGlnAlaLeuAlaPheArgGlnAspLeuGlnValAlaSerSerThrValAlaArgL	363
Dh	158486	TGGCTGAGCGCCCTAGGCTTTCATGTATCTGAGAGGTGGGCCAACTCCACCGCGGGGCGC	158545







Oy	825	-----		825
Dp	162920	ACCCCTCCTCCATTACTGCTCACTCACCCCTCCTCCCATTCACCCCTCTCTAT		162979
Oy	825	-----		825
Dp	162980	TCACTCGTCCGACTACACCCCTCTCTCCATTTCTCTTGACACTGGCTAACCCAG		163039
Oy	825	-----		825
Dp	163040	GCCCCCTCTCCGCCCTGCCTGTCTCTCTGCGTGGACACTTCCCAGCTCTCTC		163099
Oy	826	-----AspleuSerPheaspasnSeraSpheuValme		836
Dp	163100	TGAAGTAGACCCGCTCTCTCCACAGTAGACTCTCTTGAAACAAGTAAGTCTGGTAT		163159
Oy	836	tLeuLySerleuLeuAlaIyleuSerleuProSerArgaspaSpargrThraspargl		856
Dp	163160	GTTGAGACTCTTACTGTGGCAGGCGCTACGCCAGCCGCGGATGGAGAAGTACCAAAG		163219
Oy	856	yLeuAspgInluJugluJlu-----		863
Dp	163220	CCTGGATGAGGAAGGCCAAGG-TAAGCAGACAGTGACTCTTGGGGGGGCTTGAGGA		163278
Oy	863	-----		863
Dp	163279	GGGTCTGGTGGGNGTGGCTGAGCCCTGAAGCTTCACTGTCTGTGTGTGTGTGTGT		163338
Oy	863	-----		863
Dp	163339	GTTGATGGACAGCTGCCTGTACGCGCAGGTCGCCATGACCAGCTGACTGAACAGACCG		163398
Oy	863	-----		863
Dp	163399	CCTCATGGAAATGCTTGGTGTCTGTCTGTACAGAACTTACCTTCTGGATGTTCCCGTGTCT		163458
Oy	863	-----		863
Dp	163459	GCCCCGGCAGGTTGGTCTTAAAGTCAAGGCTTACACCTACTCACTTGGCTGATGGTG		163518
Oy	863	-----		863
Dp	163519	CTCATGCTCCACAGGTGGAGCCTACCTTTGCTGCTCTGTTGGCCAGATTGACTGTAG		163578
Oy	863	-----		863
Dp	163579	TCTGGGCTGCTGAAGCAGACATGAGAGTCTTCAATAGTGGACATGCTGCACAGACAC		163638
Oy	863	-----		863
Dp	163639	AGCCTCCCATGTGAGTGGACACTTGTGGCAGAGGCGAAGATGCACCTCTAACTGG		163698
Oy	863	-----		863
Dp	163699	CCCGCATCTGACTGACCAACCTCCAGTCTGACCAACACATATACTAAGATAGTGTGGC		163758
Oy	863	-----		863
Dp	163759	ACCCGTGATAGTGGCAGATGTGCTCTGGAGTGGACATGCTTCTTCATGAGCTCGAG		163818
Oy	864	-----GI		864
Dp	163819	GTCGCGCTGGGCCCTCAGTGTTCAGCCATCACTGACCCCTGACTAACTTTCACAGATGA		163878
Oy	864	userSerialaIylSerleuProLeuAlaIyleuSerleuPherThProleuthrAlaAl		884
Dp	163879	GCCTTGGCGCGCTCCCTACACCCCTGGAGTGTCTCCCTCCACCCCACTGACTGTAG		163938
Oy	884	aGlMeAlaProlyrMetLysatgLeuSeraIrglYnThValglu-----		900
Dp	163939	CGACGTGGCTCTCTACATGAAGAGGCTGTCCCGGGGCGCTTGAAGGTGAAGTGAAGC		163998

[illegible]

```

|||||
Db 165078 GTCCTGTCAGACAGACCTTCCTTGCGGCGTACGCGCAGATTGACACAGCGCCAG 165137
OY 1016 IISerGluAlaLeuArgIleLeuHisMetGluAlaValMet 1029
Db 165138 ATATCAGAGCGCCTGGAAGATCTTACACATGAGAGGCTGTGATG 165179

RESULT 10
HSM802043 1141 bp mRNA linear PRI 18-FEB-2000
LOCUS Homo sapiens mRNA: cDNA DKFp434C0126. (from clone DKFp434C0126);
DEFINITION partial cds.
ACCESSION AL137358
VERSION AL137358.1 GI:6807876
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1141)
AUTHORS Ansoorge, W., Winkler, U., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the CDNA sequencing consortium of the
German Genome Project.
This clone (DKFp434C0126) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
Location/Qualifiers
source 1. 1141
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFp434C0126"
/cisue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1. 934
/gene="DKFp434C0126"
<1. 934
/gene="DKFp434C0126"
/note="DKFp584J0619, partial"
/codon_start=2
/product="hypothetical protein"
/protein_id="CA870710.1"
/db_xref="GI:6807877"
/translacion="LLOPHVFRSEHOGALMDCILSRILRLNKRKSRHIAAFINKEV
OFILKITYNAPALISFLQKADPDLSEFNDSNDKYLMSLLAVLSLPRDDDTGL
DEGEESGASGSLPIVSVLFTPLTAEMAPKRLSROTVGLDLEVLSDIDEMSR
REILSFSTNQRKMSAECCRNALFCLARSMONSISIAAEFLPTMYCIGSODF
EYVOTALNLPETALCOEHAVALLRATRGGVHVRDGFQADLRKPEDPAVGGRDVS
LWQPTPLQAPARVPDPRKQAEAWALLVCPRR"
1126
polyA_site 209 a 384 c 339 g 209 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 1.69e-75 Length: 1141
Score: 1359.00 Matches: 272
Percent Similarity: 98.91% Conservative: 0
Best Local Similarity: 98.91% Mismatches: 2
Query Match: 25.93% Indels: 1
DB: 9 Gaps: 0
US-09-929-769-7 (1-1029) x HSM802043 (1-1141)
OY 756 LeuLeuGlnProHisValPheArgSerGluHisGlnGlnValAlaLeuTrpAspCysLeuLeu 775

```

```

|||||
Db 2 CTGCTGACGCGACCGCTGTCCGACGACACAGCGGCGGCTGTGCGACTGCTCTG 61
OY 776 SerPheIleArgLeuLeuAsnTyrArgLysSerSerArgHisLeuAlaIlePheIle 795
Db 62 TCCCTCATCCGCGCTGCTGCTGAATTCAGGAAGCTCCCGCATTCGGCTGCTTATC 121
OY 796 AsnLysPheValGlnPheIleHisLysTyrIleThrTyrAsnAlaProAlaIleSer 815
Db 122 AACAGATTGTGACAGTTCATTCATTAAGTACATTACCTACCAATGCCAGCAGCATCTCC 181
OY 816 PheLeuGlnLysHisAlaAspProLeuHisAspLeuSerPheAspAsnSerAspLeuVal 835
Db 182 TTCCTGACAGACGACGCGCGCTCCAGACCTGCTTCACCAACACTGACCTGCTG 241
OY 836 MetLeuLysSerLeuLeuAlaGlyLeuSerLeuProSerArgAspAspArgThrAspArg 855
Db 242 ATGCTGAATTCCTCTCTTCCAGTGTCTAGCTGCTCCACGACGACGACGACGACGAC 301
OY 856 GlyLeuAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 875
Db 302 GGCCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 361
OY 876 SerLeuPheThrProLeuThrAlaAlaGlnMetAlaProTyrMetLysArgLeuSerArg 895
Db 362 TCCCTGTTACACCTCTGACCGCGCGGAGATGCGCCCTCATGAAACGCTTCCCGG 421
OY 896 GlyGlnThrValGlnAspLeuLeuGlnValLeuSerAspIleAspGlnMetSerArgArg 915
Db 422 GCCCAACGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 481
OY 916 ArgProGlnLeuLeuSerPhePheSerThrAsnLeuGlnArgLeuMetSerSerAlaGln 935
Db 482 AGACCCGACGATCTGACCTTCTCTGACCAACCTGACGCGGCTGATGACCTGCGCGAG 541
OY 936 GluGlyCysArgAsnLeuAlaPheSerLeuAlaLeuArgSerMetGlnAsnSerProSer 955
Db 542 GAGGTGTCGCGACACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
OY 956 IleAlaIleAlaPheLeuProThrPheMetTyrCysLeuGlnGlnGlnGlnGlnGlnGln 975
Db 602 ATTCACACGCTTCTCTGCGCCACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 661
OY 976 ValGlnThrAlaLeuArgAsnLeuProGlnTyrAlaLeuLeuGlnGlnGlnGlnGlnGln 995
Db 662 GTGACAGCGCGCTTCCGCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721
OY 996 ValLeuLeuHisArgAla-PheLeuValGlyMetTyrGlyGlnMetAspProSerArgI 1015
Db 722 GTGCTGCTCCACGCGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781
OY 1015 IISerGluAlaLeuArgIleLeuHisMetGluAlaValMet 1029
Db 782 GATCTCGAGCGCCTGAGATCTGCTATGAGAGCGCGGTATG 824

RESULT 11
AC020509/c 157851 bp DNA linear HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC020509
VERSION AC020509.1 GI:6664388
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 157851)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA.

```

## COMMENT

This sequence was identified as CDM:10213488 by the submitter.  
For more information on this record e-mail to fly@celera.com.

\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

## FEATURES

source

1. 157851  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

BASE COUNT 41166 a 36910 c 37402 g 42373 t  
ORIGIN

## Alignment Scores:

	Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Percent Similarity:	5.53e-42	865.00	157851	288	189	400	222	34
Best Local Similarity:	43.408	16.218						
Query Match:	16.50%							

US-09-929-769-7 (1-1029) x AC020509 (1-157851)

```

OY 1 MetHisIleLeuValAlaHisAlaMetValIleLeuLeuThrLeuGlyProProlArgAla 20
Db 82377 CTCACATATAACATCATCTACCGCTTCATTATCTGCTCACCATTATCC-----AACAGC 82324

OY 21 AspAspSerGluPheGlnAlaLeuLeuAspIleTrpPheProGlnGluLysProLeuPro 40
Db 82323 AACATGCCGCGAATCGATTACCCATTTTGACTATTGGTTCGCCCGGACGACGACGCGCC 82264

OY 41 ThrAlaPheLeuValAspThrSerGlnGluAlaLeu---LeuLeuProAspTrpLeuLys 59
Db 82263 GTCGCTTTCTGCGCCGACATCGCCGACGAGACAGTGCACACTGCTGCCGCTGCAAG 82204

OY 60 LeuArgMetIleArgSerGluValLeuArgLeuValAspAlaAlaLeuGlnAspLeuGlu 79
Db 82203 CTGAAATGATGATCGCTCATCGGTGGAGACAGCTAATTGAGGACAGCTCTCAACGATCTAACG 82144

OY 80 ProGlnGlnLeuLeuLeuPheValGlnSerPheGlyIleProValSerSerMetSerLys 99
Db 82143 CCGGATCATGCTGCTCTTTGTGCAAACTTTGGACACGCGGCTCAACTCGATGTCCAAG 82084

OY 100 LeuLeuGlnPheLeuAspGlnAlaValAlaHisAspProGlnThrLeuGlnGlnAsnIle 119
Db 82083 CTCCTGCGCATATGCTGACACACCGCTGTACTGAGACACTTTGATGTGGTGAAGATGGCATT 82024

OY 120 MetAspLysAsnTrpMetAlaHisLeuValGluValGlnHisGluArgGlyAlaSerGly 139
Db 82023 CTGAACAAGGCGCTATCTGCGCCAGCTAATCGAATTCAGACAGCGCGGCTCTAAGAAT 81964

OY 140 GlyGln---ThrPheHisSerLeuLeuThrAlaSerLeuProProlArgAspSerThr 158
Db 81963 GGGCACTACACCGCTACAGCGCTTGATGTCGATTC----- 81928

OY 159 GluAlaProLysProLysSerSerProGlnGlnProIleGlyGlnLysArgIleArgVal 178
Db 81927 -----CACTCGCAAGACTGTGCCAGATCTGCC-----AAGATCACTGTG 81889

OY 179 GlyThrGlnLeuArgVal-----LeuGlyProGluLysAspLeuAla 192
Db 81888 GTTATTACGAGAGCGCTTAATTTGATGATTACGATTCTTCACAGCTCGATGATAGACCC 81829

OY 193 GlyMetPheLeu-----GlnIlePheProLeuSerProAspProArg 206
Db 81828 ACTAATCTTCTGGCCACCAAGAGGAGTAGCCCAATCTCTCAAGCAGCCGCG----- 81775

OY 207 TrpGlnSerSerProArgProValAlaLeuAlaLeuGlnGlnAlaLeuGlnGlyGln 226
Db 81774 ---CAATTACAGTAGTCGGGAGTGAATCGCCATCTTATTTCAAAAACGTGTGCAATC 81718

OY 227 LeuAlaArgValAlaGlnLysSerProGluValProGlyIleThrValArgValLeuGln 246
Db ----- 80899

```

```

Db 81717 CTAGCAAGCCCGCAACGATAGAGGATGTGGTTAACGCCATTAACAGAGTCTAGCC 81658
OY 247 AlaLeuAlaThrLeuLeuSerSerProHisGlyAlaLeuValMetSerMetHisArg 266
Db 81657 GTG----- 81628

OY 267 SerHisPheLeuAlaCysProLeuLeuArgGlnLeuCysGlnTrpGlnArgCysValPro 286
Db 81627 -----GCCGCAACATTTTAAAG----- 81610

OY 287 GlnAspThrGlyPheSerSerLeuPheLeuLysValLeuLeuGlnMetLeuGlnTrpLeu 306
Db 81609 -----ACCTTCTTCAGCTGATCGTGCACAGGACAGATATCATCTAGTACGAG----- 81562

OY 307 AspSerProGlyValGlnGlyGlyProLeuArgAlaGlnLeuArgMetLeuAlaSerGln 326
Db 81561 -----AACGCTTCCAAAGAACTAAGTATGTTT----- 81532

OY 327 AlaSerAlaGlyArgArgLeuSerAspValArgGlyGlyLeuLeuArgLeuAlaGluAla 346
Db 81531 -----AAGCACATTCGCGCGACTCCAGTCTGCTGCAGAAATCCGAA----- 81490

OY 347 LeuAlaPheArgGlnAspLeuGluValValSerSerThrValArgAlaValIleAlaThr 366
Db 81489 ---CTCTATCATGAGAGCTTGCTTCATGCTGAGAACTCTGCGAGATTTATGCG--- 81436

OY 367 LeuArgSerGlyGlnGlnCysSerValGluProAspLeuLeuSer-----LysValLeu 384
Db 81435 -----CAGCAGTTTAAAGCGAATACCGCTTGGTGGACGAAACGAGATCGTC 81388

OY 385 GlnGlyLeuIleGlyValArgSerProHisLeuGlnGluLeuLeuThrAlaPhePheSer 404
Db 81387 CGGCGCATCTCCAA-----ACCTTCGAT 81364

OY 405 AlaThrAlaAspAlaAlaSerProPheProAlaCysLysProValValValValSerSer 424
Db 81363 CAGACCAAGGACAGC-----AAGACCGCTGCCCAAGTCCCAAGAGC 81325

OY 425 LeuLeuLeuGlnGlnGlnGluProLeuAlaGlyGlyLysProGlyAlaAspGlyLysSer 444
Db 81324 GACCACCTCTCCAC----- 81310

OY 445 LeuGluAlaValArgLeuGlyProSerSerGlyLeuLeuValAspTrpLeuGlnMetLeu 464
Db 81309 -----AAGCGCTCTTCATGCACTGCGCTCGCAATG 81277

OY 465 AspProGluValValSerSerCysProAspLeuGlnLeu-----ArgLeuLeuPhe 481
Db 81276 GATCCGAGATAGTTTCCACT-----CACTAATGAAAGAGCGCTTCTGTGTTG 81229

OY 482 SerArgArgLysGlyLysGlnAlaGlnValProSerPheArgProTyrtLeuLeuThr 501
Db 81228 TCNAAG-----TCTGACAGGAGTTTGGTTCTATCTCTGTGTC 81190

OY 502 LeuPheThrHisGlnSerSerTrpProThrLeuHisGlnCysIleLeuArgValLeuLeuGly 521
Db 81189 CTGATCAACCAACCAACCACTGAGACAGATCGAAGG---ATGCGCGAGTACTGTTC 81133

OY 522 LysSerArgGlnGlnArgPheAspProSerAlaSerLeuAspPheLeuTrpAlaCysIle 541
Db 81132 AAGAAATTTCCATGAACATGAGCTAGCCACCGCTCTCAACATCTTCGAGCACTGAC 81073

OY 542 HisValProArgIleTrpGlnGlyArgAspGlnArgThrProGlnLysArgArgGlnGlu 561
Db 81072 ACCAATCCGAAGCTGTGGAAAGGACCGGCAATATGTCGAAGAAGTACGCGCGAT 81013

OY 562 LeuValLeuArgValGlnGlyProGluLeuIleSerLeuValGluLeuIle----- 578
Db 81012 GCCTTCTCATCTGTAGAGACTCTGAACCTGAGCCGCTTCCCACTTATCTTCACGAA 80953

OY 579 ---LeuAlaGluAlaGlnTrpArgSerGlnAspGlyAspThrAlaAlaCysSerLeuIle 597
Db 80952 GGGCTTTCGAGGTCAAGCTGACAGCACTGAATAATGATTCAAGCTCTGCTCG----- 80899

```

QY 598 GlnAlaArgLeuProLeuLeuLeuSerCysCysGlyAspArgLeuSerValArgLys 617  
 Db 80898 -----CGAATGAACCTACTGTTTAACTCAGAGAGAGAGACAGACATCATGTCAG 80845  
 QY 618 ValThrGlnHisLeuSerGlyCysIleGlnGlnTrpGlyLysPheValLeuGlyArgArg 637  
 Db 80844 GTATGAGACAGCTGGAG-----AAAAGTTCGGTGTCCGACTACTTG 80803  
 QY 638 CysArgAspLeuLeuLeuGlnLeuTrpLeuGlnArgProGluLeuValProValPro 657  
 Db 80802 AAACGTGAGGTCTCCAGAGATGATCATGATATCCGGGACATCAAGTTTCAAGACCG 80743  
 QY 658 -----GluValLeuLeuHisSerGlyValAlaAlaSerSer 670  
 Db 80742 GGCAAAACCGCGAGAGCGGTACAAATTCAGAAATCTGAGAGCG----- 80698  
 QY 671 ValCysArgLeuAspGlyLeuLeuHisArgPheLeuThrLeuLeuAlaSerPheThrSer 690  
 Db 80697 -----TGGCAGCGCGACAAAGGTGTCCACAAATCTGATCACTGCTGGAGCTGTGGCC 80641  
 QY 691 SerArgAlaLeuGlnAspArgGlyAlaAlaAspAlaSerMetAlaCysArgLysLeuAlaVal 710  
 Db 80640 AAAAAGAGCTTTGAAACCTGTGTCCAGACAGAGCTGCTTCCGACAGCTGCTGCC 80581  
 QY 711 AlaHisProLeuLeuLeuLeuArgHisLeuProMetIleAlaAlaLeuLeuHisGlyArg 730  
 Db 80580 TCCACACCGCTCTCTCTCTCTCTGCGACGTGCGCTGCTGTCTGCTCATTTATGCAAGCTCG 80521  
 QY 731 ThrHisLeuAspPheGlnGlnTrpPheArgGlnGlnAsnHisLeuSerCysPheLeuHisVal 750  
 Db 80520 GCCCAGCTGACGATGAAGAGCCCTCCGCGAGAGACACATCCACGATTTGTACAGATC 80461  
 QY 751 LeuGlyLeuLeuGlnLeuLeuGlnProHisValPhe-----ArgSerGlnHis 766  
 Db 80460 CTGAGGAGCTGAGCTGTGTGCAACCAACCTTTGAGAGGCGGTACAGAACAGATC 80401  
 QY 767 GlnGlyAlaLeuTrpArgCysLeuLeuSerPheLeuArgLeuLeuLeuAsnTrpArgLys 786  
 Db 80400 CAAAACAGCTG-----TCGTCTACTTCACTCTTCAAG-----CACACAGC 80356  
 QY 787 SerSerArgHisLeuAlaAlaPheHisAsnLysPheValGlnPheIleHisLysTrpIle 806  
 Db 80355 AACGTAAAGAGCGCTCCCAATCTGACACAGATTTGTGCGAGTCTCCAGCTTACATC 80296  
 QY 807 ThrTyrsAlaProAlaAlaIleSerPheLeuGlnLysHisAlaAspProLeuHisAsp 826  
 Db 80295 AACATACATCCCTCAAGTCTCTCTCTCATGACAGATGAGGCGCATTCCTGAAGAA 80236  
 QY 827 LeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeuSerLeu 846  
 Db 80235 CTTCACACCAAGTACACCTCTACTAGGCAAACTGAGGTTTGTGTCAGGCCGTGCTG 80176  
 QY 847 ProSerArgAspArgPheThrAspArgGlyLeuAspIleGlnGly----- 861  
 Db 80175 CTGAGACCAAGTGTGCACTGGCGACGGAATGACACAGAGAAATTAACTAGCAGTAC 80116  
 QY 861 ----- 861  
 Db 80115 GATCTGATGAGCATTTGATGTAAAGCCATCGGCGACAAAGCCGTTGTAACAGAGAT 80056  
 QY 862 -----GluGlnGlnSerSerAlaGlySerLeuPro 871  
 Db 80055 CCATCGAAGTGAATCCGCAAAACCCATGATCCAGACAGCACTAGGCTTCCCTTATCG 79996  
 QY 872 LeuValSerValSerLeuPheThrProLeuThrAlaAlaGlnMetAlaProTyr----- 889  
 Db 79995 GTTCTTAACCTTTGGCTGTACAGCGCGTCAAGTACACGAGCATATCGCGCACTTCTC 79936  
 QY 890 -----MetLysArgLeuSerArgGlyGlnThrValGlnAspLeuLeuGlnValLeuSer 907  
 Db 79935 GATCTGTCAAGATCAAAAGCAAGTCAACAGAGAGACGTCGTCTTGGGTCCATCAG 79876

QY 908 AspIleAspLeuSerArgArgArgProGluIleLeuSerPhePheSerThrAsnLeu 927  
 Db 79875 GAGCTGAGAGCCCTCACTTCCAGAGATTTGTGTTCATCAACAGCTGTTCAGACAGCTG 79816  
 QY 928 GlnArgLeuMetSerSerAlaGlnGlnCysCysArgAsnLeuAlaPheSerLeuAlaLeu 947  
 Db 79815 CTCACCTTATATTTCTGCGGAGTGGCCAGATCCGGTCCATCGCTTCATCATCTGATC 79756  
 QY 948 ArgSerMetGlnAsnSerPro-----SerIleAlaAlaPheLeuProThrPhe 964  
 Db 79755 AGGATCTGAAGACAGATCCCGGCAATTCGAGACATCAACTGTGACCCCTTAAACGCTAC 79696  
 QY 965 MetLysLeuLeuGlySerGlnAspPheGlnValValGlnThrAlaAlaAsnLeuPro 984  
 Db 79695 ATTCAGTCTGGCGGAGAACTCTCTGCTGCGACCGAGCGCATTTGACAAATCTGCC 79636  
 QY 985 GluTrpAlaLeuLeuCysGlnGlnIleAlaAlaValLeuLeuHisArgAlaPheLeuVal 1004  
 Db 79635 GAGATCTGCTGCTGCTGCGAGAACAGACAGATGACATCTTAAAGCTGCTTCTGCTG 79576  
 QY 1005 GlyMetLysGlyGlnMetAspProSerAlaGlnIleSerGlnAlaLeuArgLysLeu 1023  
 Db 79575 GGCTTGAAGTGTGCTGACACTGCGCACCATGATGAAAGGTTCCTCAGACTCTGA 79519

RESULT 12  
 AC004642/c 148432 bp DNA linear INV 01-MAY-1998  
 LOCUS  
 DEFINITION  
 Drosophila melanogaster DNA sequence (Pis DS00543 (D193) and DS02867 (D200)), complete sequence.

AC004642 AC004302 AC003551 AC003552 AC003553 AC003554 AC003555  
 AC003556 AC003557 AC003558 AC003559 AC003560 AC003561 AC003562  
 AC003563 AC003564 AC003565 AC003566 AC003567 AC003568  
 AC003569 AC003570 AC003571 AC003572 AC003573 AC003574 AC003575  
 AC004642.1 GI:3097827

VERSION  
 KEYWORDS  
 SOURCE  
 Drosophila melanogaster (Subclones in Sac from P1 clones DS00543 (D193) and DS02867 (D200)) DNA.

ORGANISM  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
 AUTHORS  
 Gelinkler, S.E., Aghavani, A., Arcina, T.T., Baxter, E., Blazek, R.G., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomocan, M.A., Mazda, P., Mok, M.S., Mostreli, A.R., Mostreli, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, D., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.

TITLE  
 JOURNAL  
 REFERENCES  
 AUTHORS  
 Sequencing of Drosophila chromosome 2R, region 60B1-60B10 (bases 1 to 148432)  
 2 (bases 1 to 148432)  
 Celniker, S.E., Aghavani, A., Arcina, T.T., Baxter, E., Blazek, R.G., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomocan, M.A., Mazda, P., Mok, M.S., Mostreli, A.R., Mostreli, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, D., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.

COMMENT  
 Direct Submission  
 Submitted (01-MAY-1998) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site  
 (http://fruitfly.berkeley.edu/sequence-archive.html) or send email to drosophila@mcg.lbl.gov.



Oy	598	GlNAlaAglLeuProLeuLeuLeuSerCysCysSglYAspAspGluSerValAArgLys	617
Db	48894	-----GAAATGAAACCTTCACTGTTAAAGTCAACAGAAAGCAGCAGACCTCAATGATGTAAG	48841
Oy	618	ValThrGluHisLeuSerGlyCysIleGlnGlnTrpGlyAspSerValLeuGlyArgArg	637
Db	48840	GTAATGAGACACGTGGAC-----AAAAGTTCGGTGTCCGACACTCTTG	48799
Oy	638	CysArgAspLeuLeuGlnLeuGlnLeuTyrLeuGlnArgProGluLeuArgValProValPro	657
Db	48798	AAACGTGAGCTTCCTCCACAGATGTACATCATGTATCCCGCGCATCAAGTTTCTGAACCG	48739
Oy	658	-----GluValLeuLeuHisSerGluGlyAlaAlaSerSerSer	670
Db	48738	GGCAAAACCGCGCAGCAGCGTCAACAATTTGCAGAACTTCGAAAGCC-----	48694
Oy	671	ValCysLysLeuAspGlyLeuLeuIleHisArgPheIleThrLeuLeuAlaAspThrSerAsp	690
Db	48693	---TTCGAGCGGCAAGAGTGTCCAAACATGTGATCATCTGCTGGCAGCTGGTGCGC	48637
Oy	691	SerArgAlaLeuGluLysAsnArgGlyLysAlaAspAlaSerMetAlaCysArgLysLeuAlaVal	710
Db	48636	AAAAAGACCTTGAACCCCTGTCCAGCAGACCGAGCTGCTCTTCGGCAAGCTGCTGCC	48577
Oy	711	AlaHisProLeuLeuLeuLeuLeuArgHisLeuProMetIleAlaAlaLeuLeuHisLysArg	730
Db	48576	TCCACCCGCTTCTCTCTCTTCCTCCGACGCTGGCGCGTGTCTGTCTCATTAATGCAAGCTCGG	48517
Oy	731	ThrHisLeuAsnPheGlnGluPheArgGlnGlnAsnHisLeuSerCysPheLeuHisVal	750
Db	48516	GCCGACGCTCAGATGAAGGCCCTCGCGGAGGAGCCACCTTCACCGAATTTGTACAGATC	48457
Oy	751	LeuGlyLeuLeuGluLeuLeuGlnProHisValPhe-----ArgSerGluHis	766
Db	48456	CTGAGAGCGCTCGAGCTCTGTCACCAACCATCTTGACAGAGCGGTACAGAAGCAGATC	48397
Oy	767	GlnGlyAlaLeuTyrAspPcysLeuLeuSerPheIleArgLeuLeuAsnTyrArgLys	786
Db	48396	CAAAAACAGCTG---TCGTGTACTTCAACCTTCCAG-----CACCAACAC	48352
Oy	787	SerSerArgHisLeuAlaAlaPheIleLeuHisLysPheValGlnPheIleHisLysTyrIle	806
Db	48351	AACGTAAGAAAGAGCGCTGCCAAATGCTGTGACCAAGTTTGTGCAGATGCTCCACAGCTTACATC	48292
Oy	807	ThrTyrAsnAlaProAlaAlaIleSerPheLeuGlnLysHisAlaAspProLeuHisAsp	826
Db	48291	AACATCAATCCCTCAAGTCTCTGCTCTTCATCGAAGACGTAGCTGGGACACTCTGAAGAA	48232
Oy	827	LeuSerPheAspAsnSerAspLeuValMetLeuLysSerLeuLeuAlaGlyLeuSerLeu	846
Db	48231	CTTGGAGCCCAAGTACCTCACTAGGCAACATCGCAGGTTTGTGTCAGGCGCTGCCCTG	48172
Oy	847	ProSerArgAspAspArgThrAspArgGlyLeuAspGlnGluGly-----	861
Db	48171	CTGCAGCAACAAGTCCCACTCGCGCAGCAGGAATTTGCAGACAGAGAAAGTAAGTACAGATAC	48112
Oy	861	-----	861
Db	48111	GATCTGATGAGCATTTTGATGTATAAGCCATCGCCAGCAACCCCGTGTAAACAGAGAT	48052
Oy	862	-----GluGlnGluLysSerSerAlaGlySerLeuPro	871
Db	48051	CCCATCGAAGTAATCCGCAAAACCCCATCGATCCACAGCAGCAGTAGTAGGGTCCCTTATCG	47992
Oy	872	LeuValSerValSerLeuPheThrProLeuThrAlaAlaGluMetAlaProTyr-----	889
Db	47991	GTTTCAACCTTGCGCTCTTACAGCCGTCGGAACCTACAGGACATATTCGCCGACATTCCTC	47932
Oy	890	-----MetLysArgLeuSerArgGlyGlnThrValGluAspLeuLeuGluValLeuSer	907
Db	47931	GATCTGGTCAATCATTAAGCAGATGCCAAGCAGGAGCAGGTGTCTTGGGTCCTCCATCGAG	47872

Oy	908	ASPILEAERGLUMESERDARGATGRPROGIIULEUSERPHERSERTHIASLEU	927
Dd	47871	GAGCGGAGTGGCCCTACCTTCSCAAAGATTGTGGTTCAITCAACAGACTGTGTCAAGACTG	47812
Oy	928	GLNArgLeuMetSerSerAlaIleUGlucScysArGAsnLeuAlaPheSerLeuAlaLeu	947
Dd	47811	CTCAACCTTAATAATCTGCCGCAGTGCCSAGATCCSAGATCCSGCATCAGCTTATCATCTGATGAC	47752
Oy	948	ArgSerMetGlnAsnSerPro-----SerIleAlaAlaAlaPheLeuProThrPhe	964
Dd	47751	AGGCATCTGAAGCACAAATTCGCCGGAATTCGGACATCAACCTGTGCACCTTAGCCCTAAC	47692
Oy	965	MetTyrrCysLeuGlySerGlnAsnPheCysValValGlnThrAlaLeuArgAsnLeuPro	984
Dd	47691	AATTCAGTCTCTCCGGGAGAGAGAAGCTCTGGGTGGCAGGACGGCCATTCGACAACTGGCGC	47632
Oy	985	GLUTyrAlaLeuLeuCysGlnGlnIleIleAlaValaValLeuLeuGlnIleArgAlaPheLeuVal	1004
Dd	47631	GAGATGTCGGTGTCTCTCCAGAACACACCACAAATGCAATCTTAGAAGGTGGCTTCGCGTG	47572
Oy	1005	GlyMetTyrrGlyGlnMetAspProSerAlaGlnIleSerGluAlaLeuArgIleLeu	1023
Dd	47571	GAGTTGAAGTGGTGGCTAACACTGGCCACAGATTAGAAAGTTCTCCAGACTGTA	47515

RESULT 13	LOCUS AC099018	DEFINITION Drosophila melanogaster, chromosome 2R, region 60A-60B, BAC clone BACR28M05, complete sequence.	ACCESSION AC099018	VERSION AC099018.1	KEYWORDS GI:16798950	ORGANISM HMG.	SOURCE Drosophila melanogaster	REFERENCE 1 (bases 1 to 154840)	AUTHORS Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Fartan,D., Ferreira,S., Frisoe,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Patel,C., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.	TITLE Drosophila melanogaster, chromosome 2R, region 60A-60B	JOURNAL Unpublished	REFERENCE 2 (bases 1 to 154840)	AUTHORS Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Fartan,D., Ferreira,S., Frisoe,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Patel,C., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.	TITLE Drosophila melanogaster, chromosome 2R, region 60A-60B	JOURNAL Submitted (08-NOV-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US	COMMENT Sequence submitted by: Berkeley Drosophila Genome Project
-----------	----------------	--	--------------------	--------------------	----------------------	---------------	--------------------------------	---------------------------------	--	--	---------------------	---------------------------------	--	--	--	---









Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
 Nusskern,D.R., Pacleb,J.M., Palazolo,M., Piltman,G.S., Pan,S.,  
 Pollard,J., Puri,Y., Reese,M.G., Reinert,K., Remington,K.,  
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,  
 Stimpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,  
 Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,  
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Massaran,D.A.,  
 Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,  
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
 Zaveri,I.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
 The genome sequence of *Drosophila melanogaster*  
 Science 287 (5461), 2185-2195 (2000)

JOURNAL MEDLINE  
 20196006  
 TITLE  
 The genome sequence of *Drosophila melanogaster*  
 PUBMED  
 10731132  
 REFERENCE  
 2 (bases 1 to 303823)  
 AUTHORS  
 Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D.,  
 Banzon,J., Beeson,K.Y., Busan,D.A., Carlson,J.W., Center,A.,  
 Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,Y.,  
 Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferrera,S., Flise,E.,  
 Gall,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J.,  
 Hoskins,R.A., Hostin,D., Howland,T.J., Idegawa,C., Jalali,M.,  
 Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M.,  
 Murphy,B., Nelson,C., Nelson,K.A., Nunoo,D., Pacleb,J., Paragas,V.,  
 Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Piltman,G.S.,  
 Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R.,  
 Svirskas,R., Tector,C., Tyler,D., Williams,S.M., Zaveri,J.S.,  
 Smith,H.O., Venter,J.C. and Rubin,G.M.  
 Sequencing of *Drosophila melanogaster* genome  
 Unpublished  
 3 (bases 1 to 303823)  
 JOURNAL  
 REFERENCE  
 4 (bases 1 to 303823)  
 AUTHORS  
 Misra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L.,  
 Campbell,K., Hradscky,P., Huang,Y., Kaminker,J.S., Prochnik,S.E.,  
 Smith,C.D., Tupy,J.L., Bergman,C.M., Bertram,B.P., Carlson,W.,  
 Celniker,S.E., Clamp,M.E., Drysdale,R.A., Emmert,D., Frise,E., de  
 Grey,A.D.N.J., Harris,N.L., Krommiller,B., Marshall,B.,  
 Millburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E.,  
 Shu,S., Smutnick,F., Whitfield,E.J., Yamada,C., Ashburner,M.,  
 Gelbart,W.M., Rubin,G.M., Mungall,C.J. and Lewis,S.E.  
 Annotation of *Drosophila melanogaster* genome  
 Unpublished  
 4 (bases 1 to 303823)  
 JOURNAL  
 REFERENCE  
 5 (bases 1 to 303823)  
 AUTHORS  
 Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.  
 Direct Submission  
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 JOURNAL  
 TITLE  
 Direct Submission  
 COMMENT  
 Submitted (31-MAY-2002) University of California Berkeley, 539 Life  
 Sciences Addition, Berkeley, CA 94720, USA  
 On Jun 28, 2002 this sequence version replaced gi:7291637.  
 location/Qualifiers  
 1. 303823  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /map="2R"  
 /chromosome="2"  
 /map="60A9-60A10"  
 /gene="CG11169"  
 /complement(join(11873..23611))  
 /db\_xref="FLYBASE:FBan0011169"  
 /db\_xref="FLYBASE:FBgn0034905"  
 /complement(join(11873..12379,18020..18143,23421..23611))  
 /gene="CG11169"  
 /product="CG11169-RA"  
 /db\_xref="FLYBASE:FBan0011169"  
 /db\_xref="FLYBASE:FBgn0034905"  
 /complement(join(11873..12379,18020..18091))  
 /gene="CG11169"  
 /note="CG11169 gene product"  
 /codon\_start=1

/product="CG11169-PA"  
 /protein\_id="AA047060.1"  
 /db\_xref="GI:7291638"  
 /db\_xref="FLYBASE:FBan0011169"  
 /db\_xref="FLYBASE:FBgn0034905"  
 /translation="MSSPLVLFALLVSSSTICQGLPDLKQMTPEQLQSVFHYDTH  
 AVHYELVQIHHENHNRRRRSGKPKYNAIPHHVKKDLKNAVYSELHEALAA  
 SGNHLEPAPEVSAIKSHNVSFSAFGQLNLRLATQGLFKGAPHLQMLMTVGSPPNAT  
 HGIDLEIVHEQVSRVYSMTOLKIKHKRVY"  
 complement(16742..17216)  
 /gene="CG15800"  
 /map="60A10-60A10"  
 /db\_xref="FLYBASE:FBan0015800"  
 /db\_xref="FLYBASE:FBgn0034904"  
 complement(16742..17216)  
 /gene="CG15800"  
 /product="CG15800-RA"  
 /db\_xref="FLYBASE:FBan0015800"  
 /db\_xref="FLYBASE:FBgn0034904"  
 complement(16743..17216)  
 /gene="CG15800"  
 /note="CG15800 gene product"  
 /codon\_start=1  
 /product="CG15800-PA"  
 /protein\_id="AA047061.1"  
 /db\_xref="GI:7291639"  
 /db\_xref="FLYBASE:FBan0015800"  
 /db\_xref="FLYBASE:FBgn0034904"  
 /translation="WRFETRDKRTISVNLALLESKLVIREKCVQPEDEPVLP  
 HGHISNVLKILMAEYHEKHEEPAWDSDDPLAEIEIQISWDSEFLKVELDISIC  
 KIMECNVLDIPMLYKCAQKLVELSSREBETQSGYVEKIPRODHFIRDOKE"  
 27566..28291  
 /gene="CG30177"  
 /map="60A10-60A10"  
 /db\_xref="FLYBASE:FBan0030177"  
 /db\_xref="FLYBASE:FBgn0050177"  
 27566..28291  
 /gene="CG30177"  
 /product="CG30177-RA"  
 /db\_xref="FLYBASE:FBan0030177"  
 /db\_xref="FLYBASE:FBgn0050177"  
 27671..28291  
 /gene="CG30177"  
 /note="CG30177 gene product"  
 /codon\_start=1  
 /product="CG30177-PA"  
 /protein\_id="AA068270.1"  
 /db\_xref="GI:2162663"  
 /db\_xref="FLYBASE:FBan0030177"  
 /db\_xref="FLYBASE:FBgn0050177"  
 /translation="MLRLVYIYGTLLPGWLSLRAGIDORHFIELMLKVVYVALLO  
 GLGLTLPFLGLGSFYAGCKILISVGLFMSPYSTNNHFNLLNTYSNGKICPPYEKCY  
 RWHKVCNRIETRGFMKSLVTYVMPRODEHITPRTINRLKLEKSLNMTQIAKNKA  
 DRRHQHSPMIIHSPRHLRAELDKFRESMHNVIDYLGORGSDQO"  
 complement(30339..30720)  
 /gene="CG13561"  
 /map="60A11-60A11"  
 /db\_xref="FLYBASE:FBan0013561"  
 /db\_xref="FLYBASE:FBgn0034906"  
 complement(join(30339..30494,30556..30720))  
 /gene="CG13561"  
 /product="CG13561-RA"  
 /db\_xref="FLYBASE:FBan0013561"  
 /db\_xref="FLYBASE:FBgn0034906"  
 complement(join(30339..30494,30556..30720))  
 /gene="CG13561"  
 /note="CG13561 gene product"  
 /codon\_start=1  
 /product="CG13561-PA"  
 /protein\_id="AA047062.1"  
 /db\_xref="GI:7291640"  
 /db\_xref="FLYBASE:FBan0013561"  
 /db\_xref="FLYBASE:FBgn0034906"  
 /db\_xref="FLYBASE:FBgn0034906"





Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, L.,  
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Store, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Umanli, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
-----  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
-----  
2 (bases 1 to 144097)  
Direct Submission  
Submitted (05-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 144097)  
Worley, K.C.  
Direct Submission  
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 18, 2002 this sequence version replaced gi:20452937.  
-----  
Genome Center of Medicine  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
-----  
Project Information  
Center project name: GNMW  
Center clone name: CH230-101110  
-----  
Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 62007 bases at least Q40  
Consensus quality: 68326 bases at least Q30  
Consensus quality: 73396 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see <http://www.hgsc.bcm.tmc.edu/docs/Genbankdraftdata.html>).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 72 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
1 1060: contig of 1060 bp in length  
\*  
1061 1160: gap of unknown length  
\*  
1161 2270: contig of 1110 bp in length  
\*  
2271 2370: gap of unknown length  
\*  
2371 3371: contig of 1001 bp in length  
\*  
3372 3471: gap of unknown length  
\*  
3472 4834: contig of 1363 bp in length  
\*  
4835 5992: gap of unknown length  
\*  
5993 6092: contig of 1058 bp in length  
\*  
6093 7670: gap of unknown length  
\*  
7671 7770: contig of 1578 bp in length  
\*  
7771 9212: gap of unknown length  
\*  
9213 9312: contig of 1442 bp in length  
\*  
9313 10541: gap of unknown length  
\*  
10542 10641: contig of 1229 bp in length  
\*  
10642 12072: gap of unknown length  
\*  
12073 12172: contig of 1431 bp in length  
\*  
12173 13552: gap of unknown length  
\*  
13553 13652: contig of 1380 bp in length  
\*  
13653 15096: gap of unknown length  
\*  
15097 15196: contig of 1444 bp in length  
\*  
15197 16222: gap of unknown length  
\*  
16223 16322: gap of unknown length  
\*  
16323 17655: contig of 1333 bp in length  
\*  
17656 17755: gap of unknown length  
\*  
17756 18939: contig of 1184 bp in length  
\*  
18940 20566: gap of unknown length  
\*  
20567 20666: contig of 1527 bp in length  
\*  
20667 21987: gap of unknown length  
\*  
21988 22087: contig of 1321 bp in length  
\*  
22089 23119: gap of unknown length  
\*  
23120 23219: contig of 1032 bp in length  
\*  
23219 24254: gap of unknown length  
\*  
24255 24354: contig of 1035 bp in length  
\*  
24355 25397: gap of unknown length  
\*  
25398 25497: contig of 1043 bp in length  
\*  
25498 27045: gap of unknown length  
\*  
27046 27145: contig of 1548 bp in length  
\*  
27146 28578: gap of unknown length  
\*  
28579 28678: contig of 1433 bp in length  
\*  
28679 29972: gap of unknown length  
\*  
29973 30072: contig of 1234 bp in length  
\*  
30073 31567: gap of unknown length  
\*  
31568 33101: contig of 1495 bp in length  
\*  
33101 33201: gap of unknown length  
\*  
33202 34444: contig of 1434 bp in length  
\*  
34445 34544: gap of unknown length  
\*  
34545 35824: contig of 1243 bp in length  
\*  
35825 35924: gap of unknown length  
\*  
35925 37319: contig of 1280 bp in length  
\*  
37319 37418: gap of unknown length  
\*  
37419 38684: contig of 1394 bp in length  
\*  
38685 38784: gap of unknown length  
\*  
38785 40419: contig of 1266 bp in length  
\*  
40420 40519: gap of unknown length  
\*  
40520 42463: contig of 1635 bp in length  
\*  
42464 42563: gap of unknown length  
\*  
42564 44280: contig of 1944 bp in length  
\*  
44281 44380: gap of unknown length  
\*  
44381 45965: contig of 1717 bp in length  
\*  
45966 46065: gap of unknown length  
\*  
46066 47457: contig of 1585 bp in length  
\*  
47458 47557: gap of unknown length  
\*  
47559 49418: contig of 1392 bp in length  
\*  
49419 49518: gap of unknown length  
\*  
49519 50846: contig of 1861 bp in length  
\*  
50847 50946: gap of unknown length  
\*  
50947 53079: contig of 1498 bp in length  
\*  
53080 53179: gap of unknown length  
\*  
53180 54796: contig of 2133 bp in length  
\*  
54797 54896: gap of unknown length  
\*  
54897 55980: contig of 1617 bp in length  
\*  
55981 56080: gap of unknown length  
\*  
56081 58602: contig of 1084 bp in length  
\*  
58603 58702: gap of unknown length  
\*  
58703 60423: contig of 2522 bp in length  
\*  
60424 60523: gap of unknown length  
\*  
60524 62328: contig of 1721 bp in length  
\*  
62329 62428: gap of unknown length  
\*  
62430 64486: contig of 1805 bp in length  
\*  
64487 64586: gap of unknown length  
\*  
64587 66599: contig of 2058 bp in length  
\*  
66600 66759: gap of unknown length  
\*  
66760 68152: contig of 2073 bp in length  
\*  
68153 68252: gap of unknown length  
\*  
68253 70718: contig of 1393 bp in length  
\*  
70719 70818: gap of unknown length  
\*  
70819 72273: contig of 2466 bp in length  
\*  
72274 72373: gap of unknown length  
\*  
72374 74454: contig of 1455 bp in length  
\*  
74455 74555: gap of unknown length  
\*  
74556 76795: contig of 2081 bp in length  
\*  
76796 76895: gap of unknown length  
\*  
76896 16222: gap of unknown length



```

Oy      299  euEnGlmleuEnlInfrIleuAspSerProGlyValGIuGIuLyIyIyPro----- 315
      :::      |||      ::|||      |||||      |||||
Db      51228  -GTAAGAAAGCCTTGGGTAATCTCCCAAGTCAACCCCTG-----GGCCCGGCTGTTCTCC 51177
Oy      316  -----LeuArgAlaGlnLeuArgMetLeuAlaSerGlnAlaSerAlaGlyArgArgL 333
      |||      |||      |||      |||      |||
Db      51176  CAGCATGCGCTAGGCGCTCCCTCCGCGCAGGTTCTCTTAAACACAAAGCAGCTACCT 51117
Oy      333  euSerAspValArgGly-----GlyLeuLeuArgLeuAlaGlnAlaLeuAlaPheA 350
      ||      ::|||      |||      ::|      |||
Db      51116  TGGACACAGGACGAGGCGGCATGCGACAGGACAGAGTCCCATCTTAAGGT---GACACGAA 51060
Oy      350  rGgIAspLeuGluValSerSerThrValArgAlaValIleAlaThrIleuArgSerG 370
      |||      |||      ::|||      |||      ::|
Db      51059  GGAGGCTCTCTGCACCCATCTTCGATACACAGAGTCGATGTGTG----- 51016
Oy      370  LyGIuGInCysSerValGluProAspLeuIleSerIyIyAl 383
      |||      |||      ::|||      |||      |||
Db      51015  -----TGATGGTGTGTATGCTGTGTTTCAGCAGAGGTT 50983

```

```
Search completed: May 14, 2003, 01:59:16
Job time : 8257 secs
```